

UNIVERSIDAD ADOLFO IBÁÑEZ FACULTAD DE INGENIERÍA Y CIENCIAS

Maestría en Inteligencia Artificial (MIA) Avance primera parte de
Capstone

Título del trabajo:

Predicción de Demanda Hospitalaria de Insumos Médicos

Alumnos:

Juan Luis Contreras
Yohan Alfonso Hernández
Nelson Troncoso
Victor Rodriguez

23 de Noviembre 2025

Tabla de contenido

1. Resumen.....	2
2. Introducción	3
Elemento Innovador del Proyecto.....	4
3. Descripción del Problema.....	5
Contexto: Empresa Solventum	5
Instancias de venta de Productos elegidas.....	5
Analogía de venta de Solventum	6
Metodología actual de trabajo detección de oportunidades	7
4. Objetivos.....	8
Solución propuesta	9
Metodología de trabajo Propuesta con solución IA	10
5. Datos y Variables Relevantes	11
Fuente de los Datos	11
6. Flujo de Trabajo	12
Etapa 1: Ingesta y limpieza.....	12
Etapa 2: Estandarización de productos	16
Etapa 3: Análisis Exploratorio de Datos (EDA).....	21
Etapa 4: Agregación y generación de dataset de entrenamiento	46
Etapa 5: Modelado	49
Etapa 6: Exploración del modelo y recomendación	60
7. Análisis de sensibilidad y costos.....	65
8. Conclusiones y trabajos futuros.....	70
Anexo	73
Diccionario de Datos - Documentos de compras de productos médicos.....	73
Referencias bibliográficas	76

1. Resumen

El presente trabajo tiene como objetivo principal desarrollar un sistema predictivo de la demanda mensual de insumos médicos en el sistema de compras públicas de Chile, diseñado específicamente para fortalecer la estrategia comercial y de marketing de la empresa Solventum. A diferencia de los modelos logísticos tradicionales, esta herramienta busca anticipar las necesidades de consumo de las instituciones de salud para que el equipo de ventas pueda identificar y priorizar oportunidades de negocio de manera proactiva.

Para ello, se analizó el historial de adquisiciones en Mercado Público (Convenio Marco, Compra Ágil y Trato Directo), utilizando técnicas de Inteligencia Artificial para estandarizar descripciones de productos y entrenar modelos de regresión de alta precisión. El modelo no solo estima volúmenes, sino que traduce datos no estructurados en insights accionables a través de una interfaz de lenguaje natural (GenAI).

Con esta solución, Solventum podrá reducir en un 80% el tiempo dedicado al análisis manual de oportunidades, permitiendo que sus promotores enfoquen sus esfuerzos en visitas a hospitales con mayor potencial de demanda, optimizando así la conversión comercial y reforzando su posicionamiento competitivo en el sector público.

2. Introducción

En el ecosistema de salud pública chileno, el éxito de los proveedores de insumos médicos no solo depende de la calidad de sus productos, sino de su capacidad para anticipar ciclos de compra y posicionarse estratégicamente antes de que se generen las órdenes de adquisición. Para empresas como Solventum, que operan bajo un modelo de generación de demanda a través de distribuidores, la visibilidad sobre el consumo real de los hospitales es un activo crítico para orientar sus esfuerzos de marketing y promoción en terreno.

Actualmente, el proceso de detección de oportunidades en plataformas como ChileCompra es predominantemente manual, lo que consume hasta un día completo de la jornada semanal de los vendedores y genera una respuesta reactiva frente al mercado. En este contexto, la analítica predictiva y la inteligencia artificial surgen como motores de transformación comercial, permitiendo transitar de una búsqueda intuitiva a una planificación basada en evidencia.

Este proyecto desarrolla un modelo de predicción de demanda hospitalaria que permite a Solventum conocer, con granularidad de producto y hospital, qué instituciones requerirán insumos en los próximos meses. La innovación radica en la capacidad de procesar texto libre del sistema público y exponer los resultados mediante una API asistida por IA Generativa, facilitando que el equipo comercial consulte predicciones en lenguaje natural.

El propósito final no es la gestión de inventario propio, sino la optimización de la fuerza de ventas: liberar tiempo analítico para maximizar el impacto de las visitas hospitalarias y asegurar que los productos de Solventum sean la opción preferida por los clínicos en el momento preciso de la necesidad.

Elemento Innovador del Proyecto

La innovación de este proyecto no reside solo en el uso de algoritmos avanzados, sino en la creación de una capa de inteligencia comercial que traduce datos públicos caóticos en decisiones tácticas para el equipo de ventas de Solventum. Mientras que el proceso tradicional se basa en la intuición o en el análisis manual reactivo, esta solución implementa un enfoque de proactividad estratégica basado en tres ejes disruptivos:

De la Búsqueda Manual al "Targeting" de Alta Precisión: En lugar de que el vendedor dedique horas a filtrar palabras clave en Mercado Público para adivinar el consumo de un hospital, el sistema entrega una estimación numérica granular (unidades reales por producto y hospital). Esto permite transitar de una planificación basada en "corazonadas" a una basada en potencial de volumen, asegurando que el esfuerzo de promoción se concentre donde el retorno es mayor.

Tratamiento Semántico de Datos no Estructurados: Uno de los mayores desafíos en las compras públicas es la falta de estandarización en las descripciones de los ítems. La innovación técnica aquí radica en el desarrollo de un motor de estandarización semántica (usando embeddings y matching avanzado) capaz de identificar productos de Solventum y de la competencia dentro de miles de registros de texto libre. Esto otorga una visibilidad del mercado que antes era técnicamente imposible de procesar a gran escala.

Democratización de la Analítica mediante GenAI: La solución culmina en un interfaz conversacional (Co-piloto de Ventas) que elimina la barrera entre el dato complejo y el usuario comercial. El vendedor no interactúa con tablas de Excel, sino con un asistente que entiende consultas en lenguaje natural (ej. “¿Cuál es el hospital con mayor demanda proyectada de apósitos para el próximo mes?”) y devuelve recomendaciones accionables basadas en el modelo predictivo.

Optimización Radical de la Jornada Laboral: Al reducir el tiempo de análisis manual en un 80%, la herramienta no solo mejora el dato, sino que rediseña la función del vendedor. Al liberar casi un día completo de trabajo administrativo a la semana, la tecnología permite que el recurso humano se enfoque en lo que realmente genera valor para Solventum: la relación clínica y la promoción técnica en el hospital.

Esta combinación de Forecasting de demanda, Procesamiento de Lenguaje Natural y Generative AI aplicada al marketing médico, representa un salto cualitativo en la forma en que las empresas del rubro interactúan con el sistema público de salud.

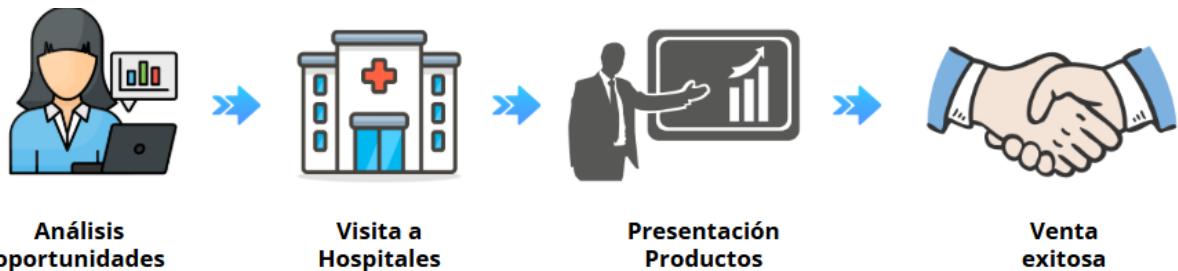
3. Descripción del Problema

Contexto: Empresa Solventum

Nos encontramos en el contexto de las compras públicas en Chile, específicamente en el sistema gestionado por ChileCompra a través del portal Mercado Público, donde los organismos del Estado realizan procesos de adquisición de bienes y servicios.

En este proyecto nos enfocamos en la compra de insumos médicos por parte de hospitales. Dentro de este ecosistema participa Solventum, una empresa que ofrece productos médicos de alta calidad.

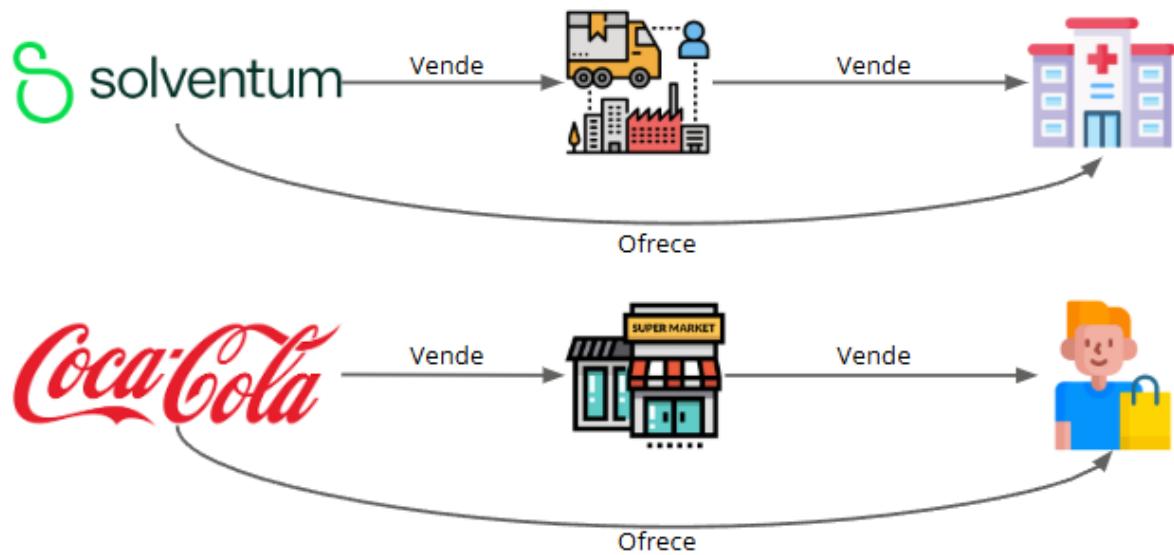
El proceso consta de lo siguiente: los vendedores de Solventum buscan oportunidades y visitan los hospitales y presentan las ventajas de sus productos para que estos sean solicitados en las instancias de compras habilitadas.



Instancias de venta de Productos elegidas

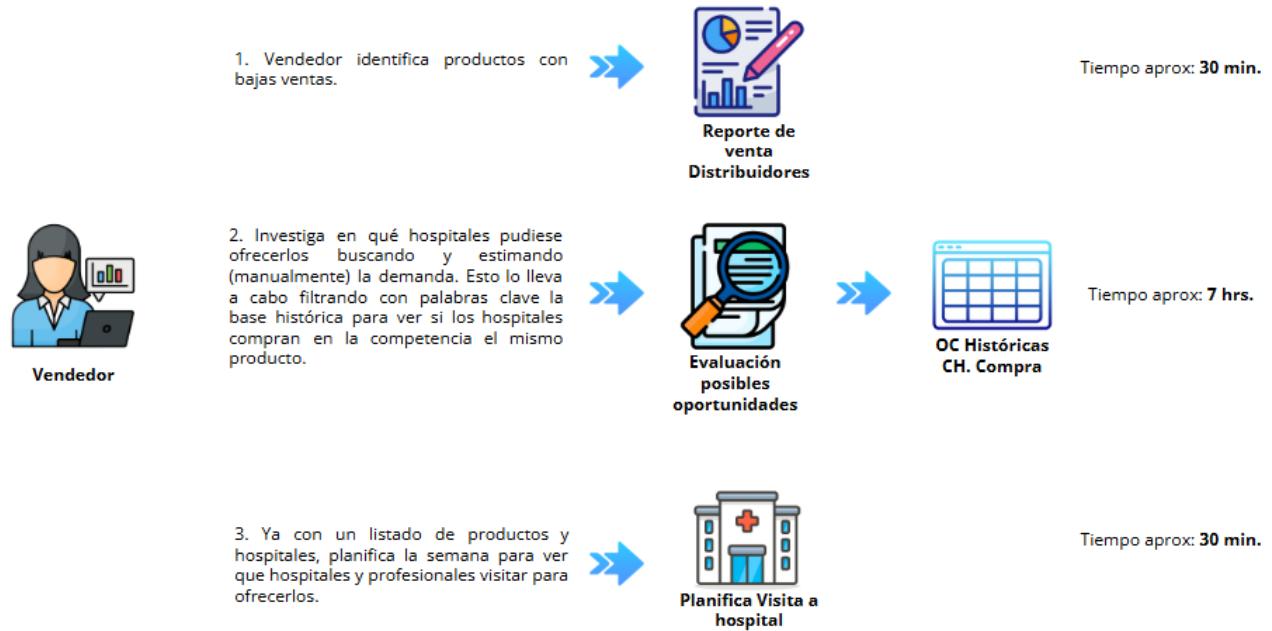
- *Trato Directo*: Es un procedimiento de compra excepcional que permite contratar con un proveedor por las causales expresamente indicadas en la normativa.
- *Convenio Marco*: Los Convenios Marco son un procedimiento de compra que se traduce en una tienda electrónica en donde los compradores públicos pueden adquirir bienes y servicios a través de un solo clic. Los rubros de cada Convenio Marco cumplen con las recomendaciones de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) respecto a representar rubros
- *Compra Ágil*: Los organismos del Estado e instituciones regidas por la Ley de Compras Públicas tienen a su disposición la Compra Ágil para realizar sus compras menores de manera expedita y simple. Con un monto máximo de 100 UTM por cada compra y solo las empresas de menor tamaño pueden participar.

Analogía de venta de Solventum



Metodología actual de trabajo detección de oportunidades

Los vendedores de Solventum tienen un objetivo primordial, propiciar la demanda de sus productos de parte de los hospitales, ellos no realizan tratos de venta directo, como se explicó previamente que son los Distribuidores quienes realmente venden a los hospitales. Sin embargo, el trabajo de marketing e incentivo ocurre igualmente. Esto se lleva a cabo visitando hospitales y entrevistando a funcionarios a quienes presentar sus productos y mostrar sus propiedades. Pero, para llegar a esto es necesario realizar un trabajo previo, con el fin de optimizar las visitas y maximizar la probabilidad de demanda de productos de Solventum. Vale mencionar que no todos los vendedores suelen realizar esta actividad, muchos, simplemente realizan trabajo en terreno sin hacer una investigación previa, sin embargo, se reporta por parte de vendedores que sí realizan esta actividad que la venta de sus productos ha aumentado un 16% anual, lo cual beneficia a Solventum y por ende obtienen mayores comisiones, es de tal manera que el proceso previo para realizar esta actividad previa es la siguiente:



4. Objetivos

Objetivo General:

Desarrollar un sistema predictivo, basado en técnicas de inteligencia artificial y expuesto mediante una interfaz conversacional (GenAI), que permita anticipar la demanda mensual de insumos médicos por institución, mejorando la planificación comercial reduciendo en 80% el tiempo de análisis manual y la identificación de oportunidades para Solventum en el sistema público de salud.

Objetivos específicos:

a) Integrar y depurar datos históricos de compras públicas

Consolidar y estandarizar la información proveniente de ChileCompra, normalizando categorías y descripciones para habilitar un análisis robusto y consistente entre hospitales logrando una tasa de similitud mínima de 90%.

b) Desarrollar modelos predictivos de alta precisión

Generar pronósticos mensuales por hospital y categoría de producto que cumplan con los estándares de calidad definidos para este proyecto:

- **MASE < 0.7** (Error absoluto medio escalado bajo el umbral de referencia)
- **sMAPE < 35%** (Error porcentual absoluto simétrico dentro de rangos aceptables para escenarios de alta variabilidad)

El cumplimiento de estas métricas busca asegurar la confiabilidad del modelo en un entorno donde la demanda es heterogénea y fluctuante.

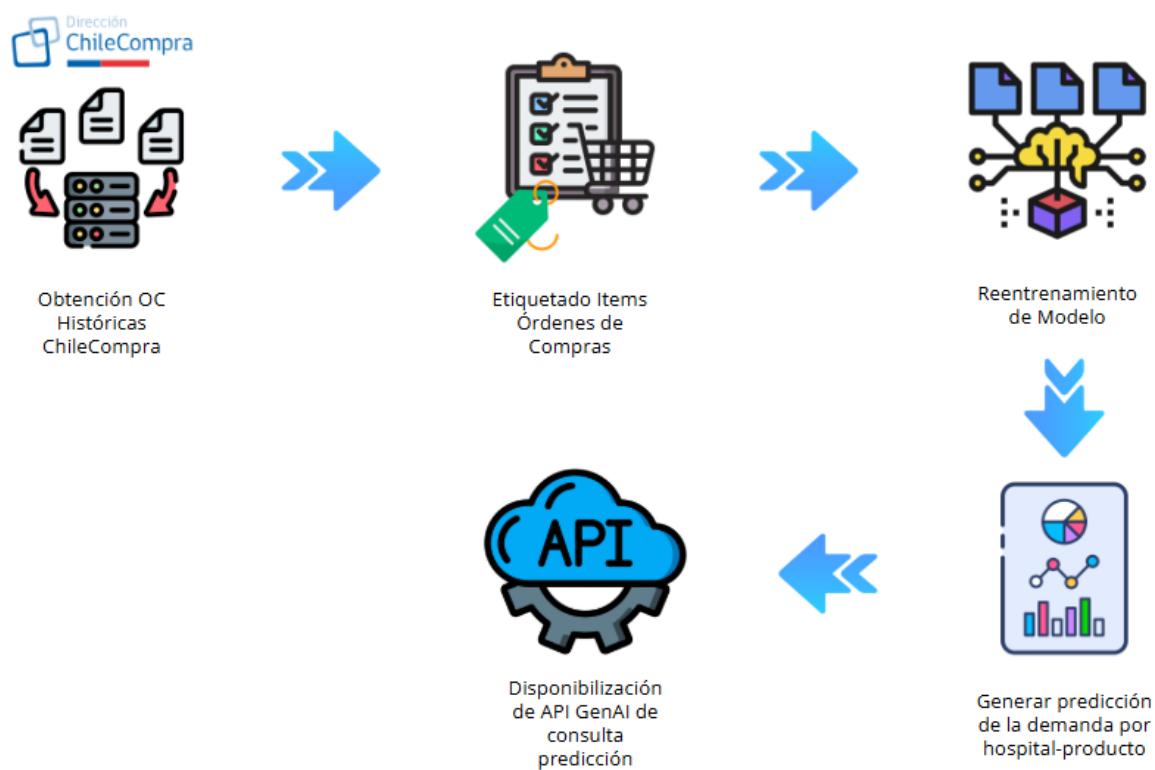
c) Exponer los resultados mediante una API asistida por GenAI

Diseñar un servicio conversacional que permita a los ejecutivos de Solventum consultar predicciones en lenguaje natural, obteniendo recomendaciones claras y accionables para apoyar la toma de decisiones comerciales

Solución propuesta

El diagrama presenta el flujo completo de la solución propuesta para predecir la demanda de insumos médicos y ponerla a disposición de los equipos comerciales mediante una API con capacidades GenAI. El proceso inicia con la obtención de órdenes de compra históricas desde ChileCompra, las cuales son posteriormente etiquetadas y estandarizadas para identificar correctamente los productos y categorías relevantes. Con esta información, se realiza el reentrenamiento periódico del modelo predictivo, que permite actualizarlo con nuevos datos y mantener su precisión.

Una vez entrenado, el modelo genera predicciones de demanda por hospital y producto, que luego se exponen mediante una API asistida por GenAI, permitiendo a los usuarios consultar las proyecciones de forma sencilla y en lenguaje natural.

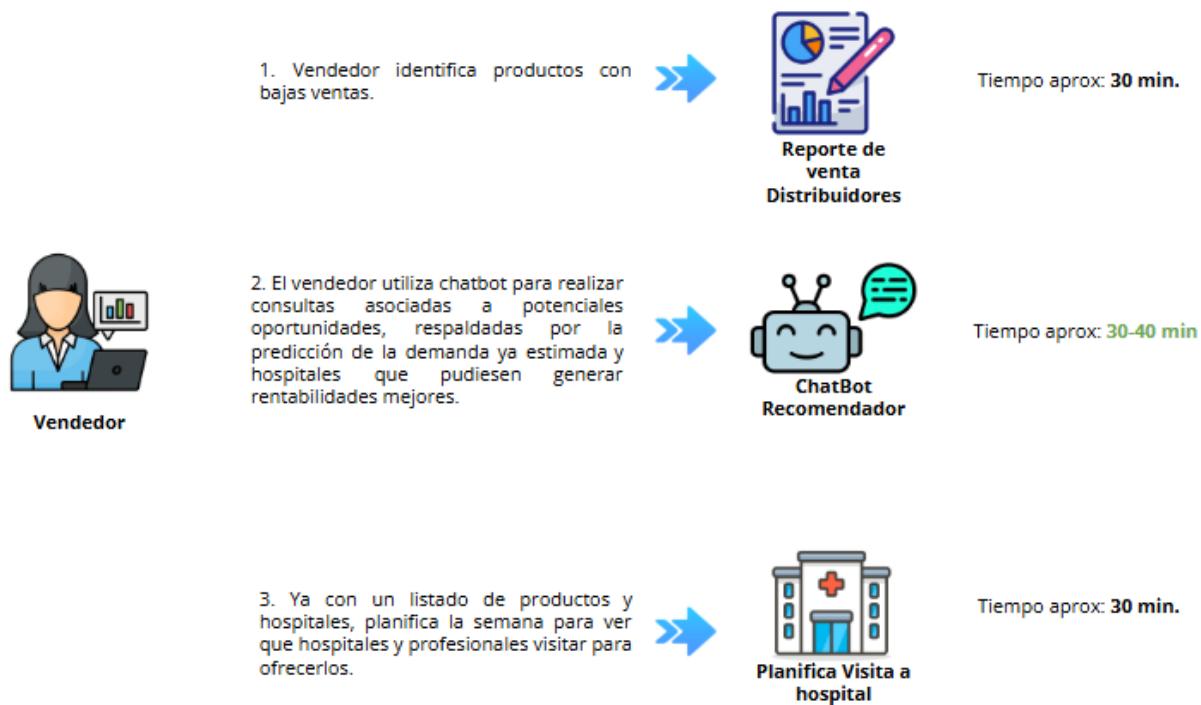


Metodología de trabajo Propuesta con solución IA

En el proceso actual, algunos vendedores dedican hasta un día completo por semana a analizar manualmente productos, ventas y potenciales oportunidades. Aunque este trabajo mejora sus resultados, reduce de forma significativa el tiempo disponible para visitas a hospitales.

La solución propuesta incorpora un Chatbot Recomendador respaldado por modelos predictivos de demanda, permitiendo obtener el mismo valor analítico —o superior— en menos de 2 horas semanales. Esto elimina la búsqueda manual de información, orienta al vendedor hacia los productos y hospitales con mayor potencial y libera tiempo para actividades de alto impacto comercial.

El diagrama resume este nuevo flujo de trabajo optimizado:



5. Datos y Variables Relevantes

Fuente de los Datos

Los datos utilizados en este proyecto provienen del **portal oficial de datos abiertos de ChileCompra**, específicamente del repositorio de órdenes y licitaciones públicas:

🔗 <https://datos-abiertos.chilecompra.cl/descargas/ordenes-y-licitaciones>

Este portal ofrece archivos detallados con el historial de órdenes de compra emitidas por instituciones del Estado de Chile. Cada registro del dataset corresponde a un **ítem individual dentro de una orden de compra**, y contiene información clave para el análisis, entre la que se destaca:

- **Identificación de la orden:** código único (OC), fecha de emisión, convenio marco, entre otros.
- **Datos del comprador:** nombre de la institución compradora, región, unidad organizacional y sector público correspondiente.
- **Información del proveedor adjudicado:** nombre, RUT, actividad económica, tamaño y región.
- **Descripción del producto:** campo de texto libre que detalla el ítem adquirido (requiere procesamiento para su estandarización).
- **Aspectos cuantitativos de la compra:** cantidad solicitada, unidad de medida, monto neto, descuentos, cargos, impuestos y monto total asociado al ítem.

Este conjunto de datos, al ser granular y extenso, permite construir series históricas ricas en patrones de comportamiento, fundamentales para el desarrollo de modelos predictivos.

Descripción resumida archivos descargados

- Contiene cientos de miles de registros.
- Las descripciones de producto se encuentran en texto libre ([DescripcionItem](#)).
- Las fechas ([FechaEnvioOC](#)) permiten agrupar por períodos.
- Las cantidades ([CantidadItem](#)) y montos ([MontoNetoItemCLP](#)) permiten modelar demanda y precios relativos.

6. Flujo de Trabajo

A continuación, se describirán las 6 etapas que se llevaron a cabo de forma secuencial, para llevar a cabo los objetivos planteados al inicio. Cada etapa depende de las anteriores y parte desde la obtención de los datos, preprocesamiento hasta finalmente la generación de un modelo predictivo.

Etapa 1: Ingesta y limpieza

Esta fase inicial tiene como objetivo preparar los datos crudos para su posterior análisis y modelado. Se enfoca en asegurar la calidad, consistencia y estructura adecuada del dataset.

Actividades realizadas:

- **Carga e integración de archivos históricos** de órdenes de compra provenientes de ChileCompra.
- **Depuración del dataset**, eliminando columnas irrelevantes o con alto porcentaje de valores nulos que no aportan valor al análisis.
- **Renombramiento y estandarización de variables** para mejorar la comprensión y facilitar su uso en etapas posteriores.
- **Corrección de tipos de datos**, asegurando que variables numéricas, fechas y categóricas estén correctamente tipificadas, y resolviendo inconsistencias como texto en columnas numéricas.
- **Exploración y caracterización de las fuentes de datos**: comprender que cada registro representa un ítem detallado dentro de una orden de compra, con atributos como hospital comprador, fecha, tipo de compra, producto, proveedor, monto, entre otros.
- **Clasificación por tipo de orden de compra**: Los obtenidos se clasifican en 4 tipos, Licitaciones, Convenio Marco, Compra Ágil y Trato Directo. Sin embargo, las Licitaciones serán descartadas, y el motivo es que ya que el objetivo es conseguir un modelo predictivo de la demanda, las Licitaciones tienen períodos fijos y largos, por lo que contar con esta información pudiera sesgar ciertos períodos y no tendrán relación con una demanda real. Dado entonces, solo trabajaremos con los últimos tres tipos de procesos de adquisición.

- **Construcción de curvas históricas:** generar series temporales de demanda agregada por hospital y por tipo de proceso de compra, para visualizar la evolución del consumo, estacionalidades, ciclos o caídas inusuales.
- **Detección de patrones y agrupamientos:** identificar instituciones con comportamientos similares, productos con alta rotación, meses críticos de demanda, y otras regularidades que puedan alimentar futuras estrategias comerciales.
- **Evaluación de consistencia y compatibilidad entre datos:** analizar si los datos de distintas modalidades de compra pueden integrarse o si es preferible construir **modelos especializados por tipo de orden**, debido a sus diferencias estructurales o de comportamiento.
- **Filtrado inteligente de registros:** definir, junto con expertos del negocio (como promotores o vendedores), qué subconjuntos de datos son más representativos del proceso de toma de decisiones y cuáles podrían introducir ruido o distorsión en el análisis.
- **Generación de nuevas variables derivadas:** construir atributos adicionales como mes y año de la compra, frecuencia de adquisiciones por hospital, precio unitario estimado, proveedor habitual, entre otros, que aportan valor explicativo a los modelos predictivos. Sin embargo las principales variables que se crearon fueron 2: Cantidad Real y Monto Real por Unidad. El primero es un derivado de la Cantidad indicada en el registro del ítem, la cual debiese ser suficiente, pero el usuario indicó que el valor presente ahí no representa la cantidad real del producto, ya que en muchas ocasiones lo que se ve ahí es la cantidad según la presentación, y este puede venir en grupos, por ejemplo cajas de 10 unidades, entonces si aparece un valor de 15 en realidad se refiere a 15 cajas de 10 unidades, lo que en realidad representa a 150 unidades reales, y eso es lo que refleja este nuevo campo. Esto se logró aplicando expresiones regulares, en donde a se extrajo la presentación indicada en la descripción y se obtiene el valor expresado ahí, luego se multiplica por la cantidad y así se obtiene la Cantidad Real. El monto Real por Unidad es simplemente aprovechar la generación del nuevo campo Cantidad real y usarlo como divisor del Monto total del ítem, por lo que si el valor total del ítem es de \$150.000 entonces el Monto real por Unidad es de \$1.000.

Descripción Item OC	Cantidad Item	Precio Unidad Producto OC	Cantidad real Item	Precio Unidad real Producto OC
1266184 APOSITO 3M TRANSPARENTE ADHESIVO TEGADERM 1626W 10 X 12 CM 50 UNIDADES	160	\$32.967	8.000	\$659
20-225-200-128-00 APOSITO TRANSPARENTE CON GASAS NO ADHERENTE 9X15 CM + PAD cJx25	10	\$31.600	250	\$1.264

- Selección de Productos:** El catálogo de Solventum presenta diferentes tipos de Insumos y Dispositivos médicos, pero para efectos de este Capstone solo seleccionaremos 2: Apóositos y Guantes Médicos. La elección de estos es de exclusiva petición del usuario de Solventum que apoya este proyecto, y la justificación es que consideran que las ventas de estos productos no son las deseadas y esperan que haya margen de crecimiento.
- Selección de Hospitales:** El listado de hospitales es amplio, sin embargo para este Capstone nos centraremos en dos: Hospital Sótero del Río y Hospital del Salvador. El motivo de la selección se debió a que, similar a lo ocurrido con los productos mencionados en el punto anterior, es que Solventum los considera clientes con alto potencial de crecimiento y suficiente historial de compra que pudiese facilitar la predicción de la demanda.

A continuación un listado de gráficos de la distribución de los productos elegidos en los hospitales seleccionados:

Apóositos:

Complejo Asistencial Dr. Sótero del Río

La serie muestra una alta variabilidad mensual en la demanda, con picos recurrentes que superan los 20.000 – 30.000 ítems y una tendencia general estable pero fluctuante a lo largo de los años.



Hospital del Salvador

La demanda presenta un comportamiento más irregular y con menor volumen, destacando algunos picos aislados y variaciones amplias entre meses, lo que sugiere una serie menos estable pero con patrones detectables.



Guantes Médicos

Complejo Asistencial Dr. Sótero del Río

La serie presenta una demanda alta y muy fluctuante, con picos que superan los 2 millones de unidades y variaciones marcadas entre meses. Se observan ciclos irregulares, pero con un volumen sostenido a lo largo del tiempo.



Hospital del Salvador

La demanda muestra **menor volumen total**, aunque con picos notorios que superan el millón de unidades. El comportamiento es **menos estable**, con oscilaciones amplias y periodos de alzas y bajas pronunciadas.



Etapa 2: Estandarización de productos

El principal desafío en esta etapa es transformar las descripciones en texto libre (**DescripcionItem**) en una codificación estandarizada que permita identificar de manera consistente productos equivalentes, especialmente aquellos relacionados con el portafolio de la empresa **Solventum**.

Descripción Item OC	Nombre Producto Estándar
(1266813) GUANTE LATEX MUNCARE NO ESTÉRIL TALLA S 100 UNIDADES 1292913	GUANTES_MEDICOS
(668385) APÓSITO 3M TRANSPARENTE ADHESIVO TEGADERM 1622W 4,4 X 4,4 CMS. CAJA DE 100 UNIDADES 668385	APOSITOS

Proceso aplicado:

- **Limpieza y normalización del texto:** eliminación de caracteres innecesarios, homogeneización de mayúsculas/minúsculas y eliminación de stopwords.
- **Tokenización y aplicación de expresiones regulares** para detectar palabras clave relevantes como “apósito”, “film”, “parche”, “transparente”, entre otras.
- **Ejemplo práctico 1:**
 - **Texto original:** (668393) APÓSITO 3M TRANSPARENTE ADHESIVO 5 X 5,7 CMS TEGADERM I.V. 1610, CAJA DE 100 UNIDADES 668393
 - **Resultado estandarizado:** APOSITOS
- **Ejemplo práctico 2:**
 - **Texto original:** (1266813) GUANTE LATEX MUNCARE NO ESTÉRIL TALLA S 100 UNIDADES 1292913
 - **Resultado estandarizado:** GUANTES_MEDICOS

Generación de identificadores de producto:

Se agrupan las descripciones similares utilizando:

- **Reglas específicas de coincidencia** basadas en los productos del catálogo Solventum.

- **Filtrado de datos** usando campos específicos de la base, como el Rubro y el código Onu, para distinguir, por ejemplo: un Guante médico con un Guante de trabajo en la construcción.
- **Coincidencia semántica** basada en embeddings de texto como **Google Ai embedings** y **Paraphrase Multilingual MiniLM L12 V2**.

Matching con productos Solventum:

El [catálogo de Solventum](#) nos permite conseguir diferentes productos que se ofrecen al mercado y que suelen estar agrupados por un tipo de producto, si bien existen variedades, ya sea por ciertas características o propiedades pero hacen referencia a uno, esto nos permite obtener varias descripciones y agruparlas en un tipo de producto específico, a continuación se presentan algunos ejemplos:

- <https://conveniomarco.mercadopublico.cl/insumos2/aposito-2228645>



- <https://conveniomarco.mercadopublico.cl/insumos2/guante-latex-controlado-2231045>



Estandarización mediante filtros y expresiones regulares

Desde acá podemos obtener palabras claves (sumando a otras proporcionadas por el usuario basándose en su propia experiencia).

De tal manera, por ejemplo para el producto “Apósitos” obtenemos lo siguiente:

- apósito, film, apósito transparente, apósito adhesivo, parche impermeable, apósito impermeable transparente, curita transparente.
- Marcas comunes de la competencia: Opsite, Leucomed.

- Ejemplo de expresión regular que pudiese potencialmente servir para filtrar:

(ap[oó]sito|film|parche)(.*?)(transparente|impermeable|adhesivo)(.*?)(opsite|leucomed)?

Resultado: Con expresiones regulares como el ejemplo anterior se logran identificar registros históricos que, aunque no mencionan directamente la marca Solventum, corresponden funcionalmente a productos similares, permitiendo estimar su demanda potencial y priorizar esfuerzos comerciales.

Estandarización mediante uso de Embeddings

Sin embargo, existe otra manera que puede ser alternativo o complementario, el cual es usar **embeddings**.

La idea es la siguiente: vectorizar las descripciones de ítems de toda la base histórica de Órdenes de Compra y también hacerlo para las descripciones de los productos del catálogo de Solventum que recién se mencionó. De tal manera que al hacer una comparación vectorial, por ejemplo, con una similitud Coseno podemos encontrar la base de Ítems de OC históricas que más se asemejan a los productos del catálogo de Solventum.

Al hacer lo anterior podemos obtener un índice de Similitud, el que nos permite escoger los registros de ítems de OCs históricas que más se asemejan a alguno de los productos de Solventum del catálogo y asignarle ese valor. Además, se debe buscar un índice de Similitud mínimo, ejemplo: 0.85 (esto se debe investigar basándonos en los resultados obtenidos).

Al realizar esta comparación obtenemos algo como lo siguiente:

AZ_desc_item_oc	AZ_desc_cat	AZ_producto	123 similitud
0110010243 APOSITO TRANSPARENTE ADHESIVO 20 X 30 CM ESTERIL	APÓSITO TEGADERM TRANSPARENTE ADHESIVO 44 X 44 CM 100 UNIDADES	APOSITOS	0.9392390251

Con el fin de encontrar ítems de OC históricos no solo de Solventum si no que también de la competencia, la generación del vector para los productos del catálogo se hizo eliminando la marca en la descripción, entiendase: 3M, Tegaderm, etc. Ya que al mantener estas palabras en la descripción, la similitud con la base histórica de las OC tenderá a preferir u obtener un índice de similitud mayor para los mismos productos de Solventum.

2.1. Validación de Calidad y Consistencia (Human-in-the-Loop)

Para garantizar que el proceso automatizado de limpieza, uso de expresiones regulares y *embeddings* sea confiable para la toma de decisiones comerciales, se implementó una metodología de validación basada en una muestra de control.

Este proceso se desglosa en los siguientes pasos:

- Selección de la Muestra: Se extrajo una muestra aleatoria representativa de 400 registros de ítems de órdenes de compra histórico.
- Criterio del Experto (Usuario): Un usuario experto de Solventum auditó manualmente cada uno de los 400 registros. Su tarea consistió en validar si el "Nombre de Producto Estándar" asignado por el modelo (ej. APOSITOS o GUANTES_MEDICOS) correspondía funcionalmente a la descripción en texto libre original.
- Cálculo de la Tasa de Acierto: La precisión del proceso de estandarización se midió utilizando la siguiente métrica:

$$\text{Accuracy} = \frac{\text{Registros Correctamente Clasificados}}{\text{Total de la Muestra (400)}} \times 100\%$$
- Resultados Obtenidos: Tras la revisión, se alcanzó una tasa de correspondencia del 95%. Esto supera el umbral del 90% definido en los objetivos específicos del proyecto, validando la robustez de la integración de datos de ChileCompra.

2.2. Refinamiento del Modelo Basado en el Feedback

El análisis de los registros que no lograron la coincidencia inicial permitió realizar ajustes críticos en el motor de estandarización:

1. Ajuste de Umbrales de Similitud: Se calibró el índice de similitud coseno (fijado inicialmente en 0.85) para evitar falsos positivos en categorías ambiguas.

2. Optimización de Expresiones Regulares: Se incorporaron nuevas palabras clave y marcas de la competencia detectadas por el usuario en la muestra, mejorando la detección semántica.
3. Filtrado de Ruido Comercial: La validación humana permitió identificar términos que el modelo confundía (ej. guantes de construcción vs. guantes quirúrgicos), permitiendo un uso más preciso del Rubro y código ONU para el descarte.

Etapa 3: Análisis Exploratorio de Datos (EDA)

Objetivo del EDA

El análisis exploratorio busca identificar patrones de comportamiento en la demanda histórica que permitan seleccionar productos, definir instituciones prioritarias y fundamentar el diseño del modelo predictivo y su aplicación comercial.

Hallazgos relevantes para la toma de decisiones

Durante el análisis se observaron tres elementos clave:

Apósitos Sótero del Río

1. Serie Temporal Original (2015–2025)
2. Serie temporal desde 2020
3. Histograma demanda apóstitos

Apóstitos Hospital El Salvador

4. Serie temporal eliminando outliers de pandemia (versión 1)
5. Serie temporal original (otra escala)
6. Serie temporal desde 2023
7. Histograma demanda apóstitos v2.

Guantes Sótero del Río

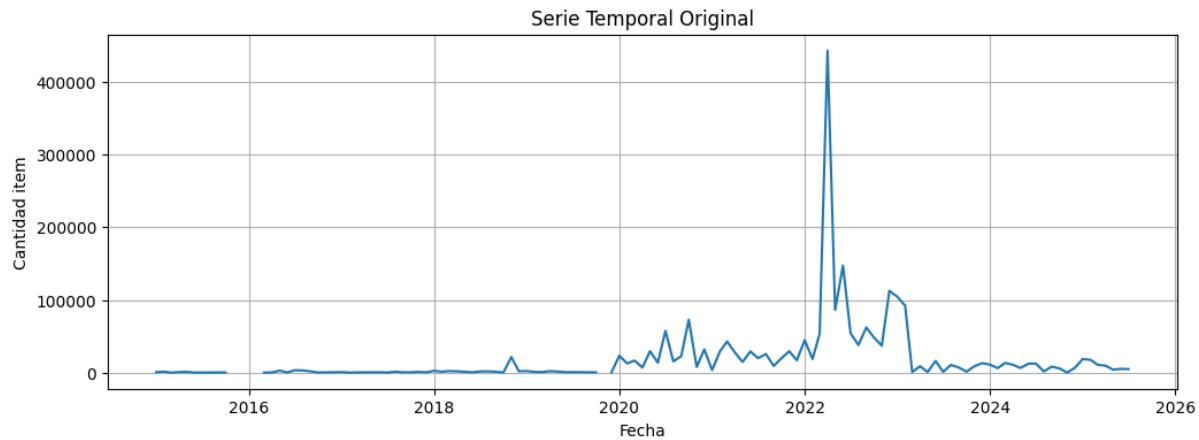
8. Serie temporal eliminando outliers (versión 2)

Guantes Hospital El Salvador

9. Serie temporal original (larga)
10. Serie temporal desde 2023 (2da versión)
11. Histograma demanda guantes
12. Serie temporal desde 2023 (guantes)

Apósitos - Sótero del Río

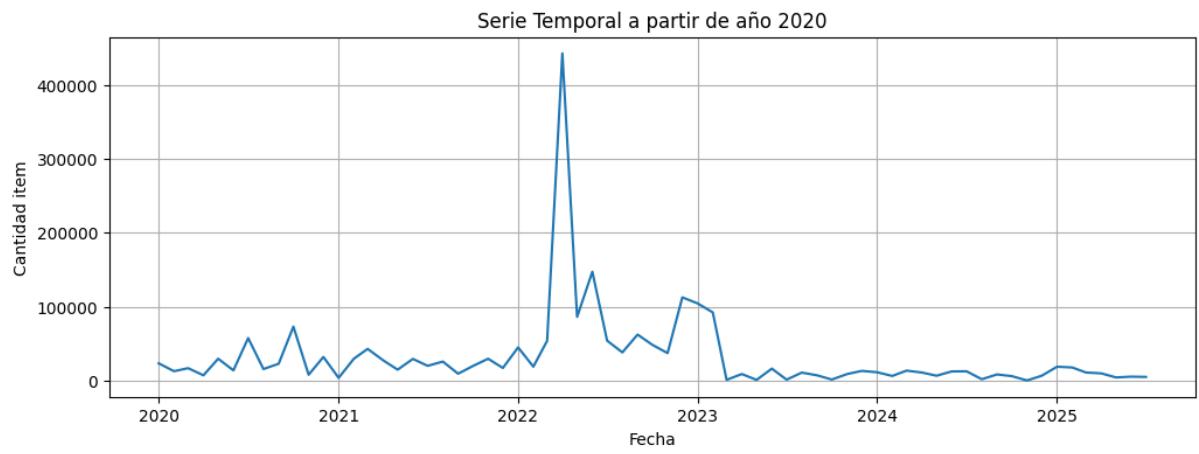
La serie inicial contiene datos desde el año 2015.



Evidencia: demanda casi nula hasta 2019; salto abrupto 2020–2022.

✓ Interpretación estratégica:

- La pandemia introduce **outliers extremos**, no representativos del comportamiento normal.
 - Estos puntos distorsionan cualquier modelo si no se corrigen.
- ✓ Conexión con forecasting: se justifica **limpieza temporal y recorte del período histórico**.

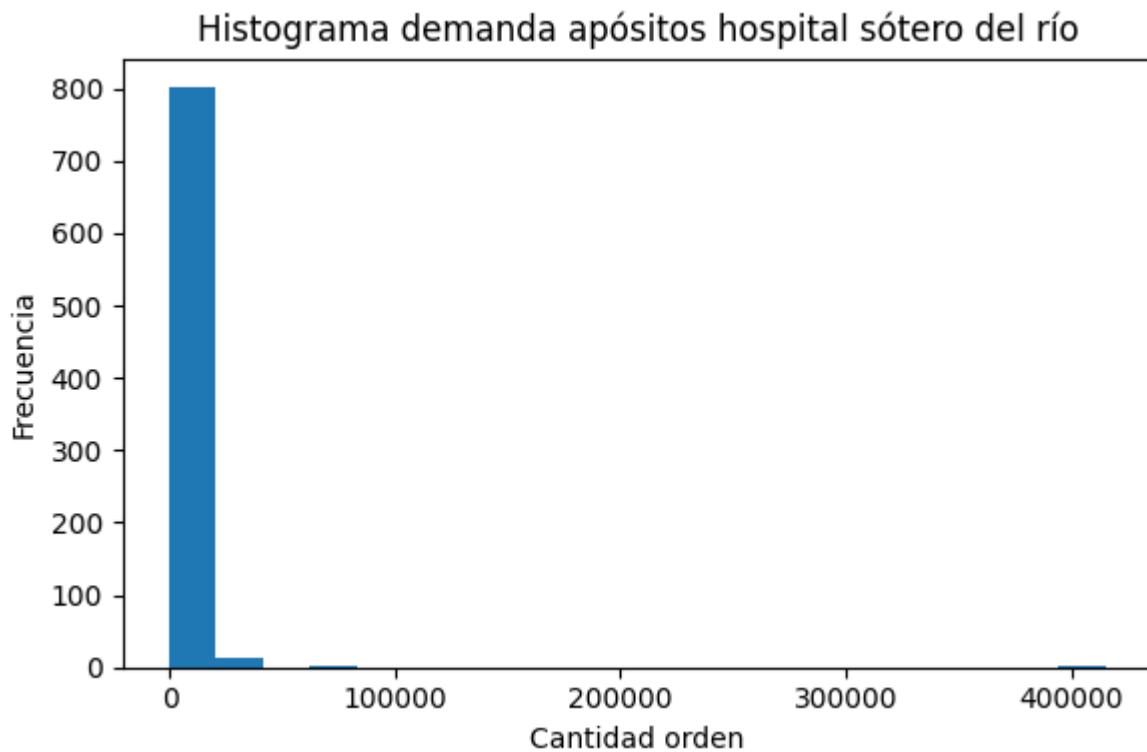


✓ Muestra la fase crítica (COVID), con picos >400.000 unidades.

✓ Interpretación:

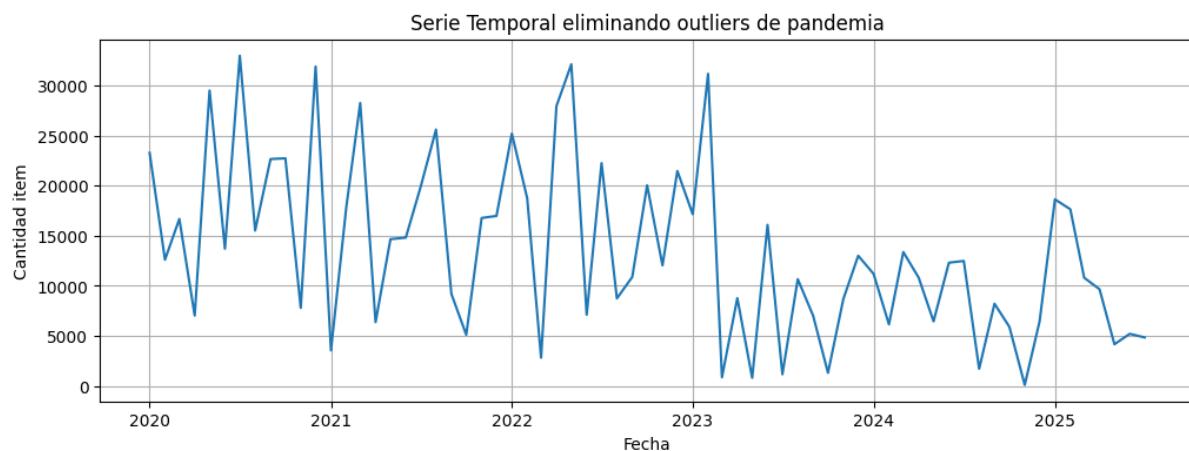
- Demanda totalmente anómala.

- No hay estacionalidad estable, solo eventos extraordinarios.
- ✓ Relevancia: confirma que el forecasting debe excluir estos años para obtener señales reales.



- ✓ Distribución extremadamente asimétrica.
- ✓ El 90% de la demanda está en valores bajos; los picos son raros pero muy grandes.
- ✓ Implicancia:
- El modelo debe manejar **series con alta dispersión y outliers**.
 - Puede justificar transformaciones (log, Box-Cox) o modelos robustos.

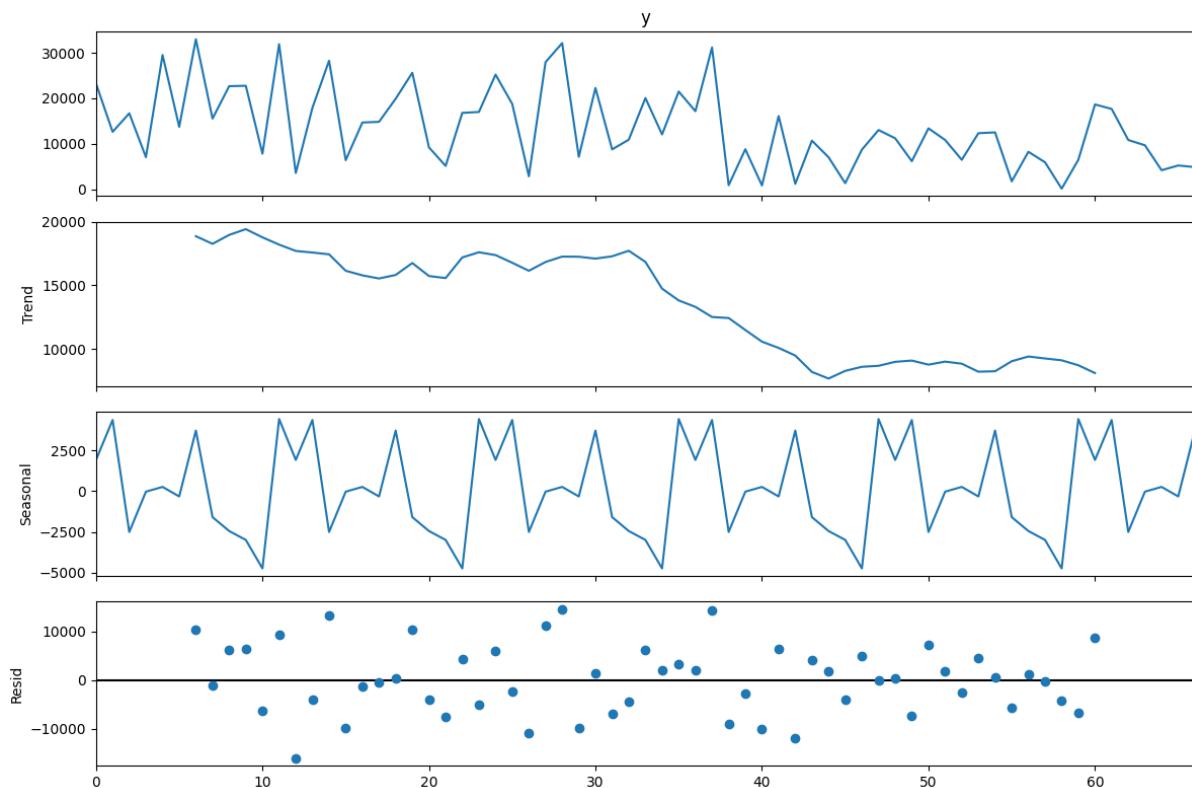
1. Descripción general de la serie



La serie observada muestra valores considerablemente volátiles, con fluctuaciones de alta magnitud a lo largo de todo el período. Visualmente se identifica un cambio progresivo en el nivel medio de la serie: en los primeros períodos los valores se ubican en un rango más elevado, mientras que en la parte final se aprecia un descenso sostenido, seguido de una leve estabilización.

Esta característica sugiere la presencia de una tendencia no constante y posibles cambios estructurales en la dinámica del proceso generador de datos.

2. Descomposición de la serie



La descomposición clásica de la serie en sus componentes de tendencia, estacionalidad y residuo permite analizar con mayor detalle su estructura interna.

2.1. Componente de tendencia

El componente de tendencia presenta un comportamiento predominantemente decreciente:

- En la primera parte del periodo se sitúa en torno a valores relativamente altos.
- A partir de un cierto punto se observa una caída progresiva y marcada.
- Hacia el final de la serie la tendencia se estabiliza en niveles más bajos, con una ligera recuperación.

Este patrón es consistente con un proceso cuya media cambia a lo largo del tiempo, por lo que la serie no es estacionaria en nivel.

2.2. Componente estacional

El componente estacional exhibe un patrón claramente periódico y estable:

- Se observan ciclos que se repiten de forma regular, con picos positivos y negativos aproximadamente simétricos.

- La amplitud de la estacionalidad se mantiene relativamente constante a lo largo del tiempo.

Lo anterior indica la existencia de una estacionalidad fuerte con un periodo fijo (por ejemplo, mensual, trimestral, etc., dependiendo de la frecuencia original de la serie).

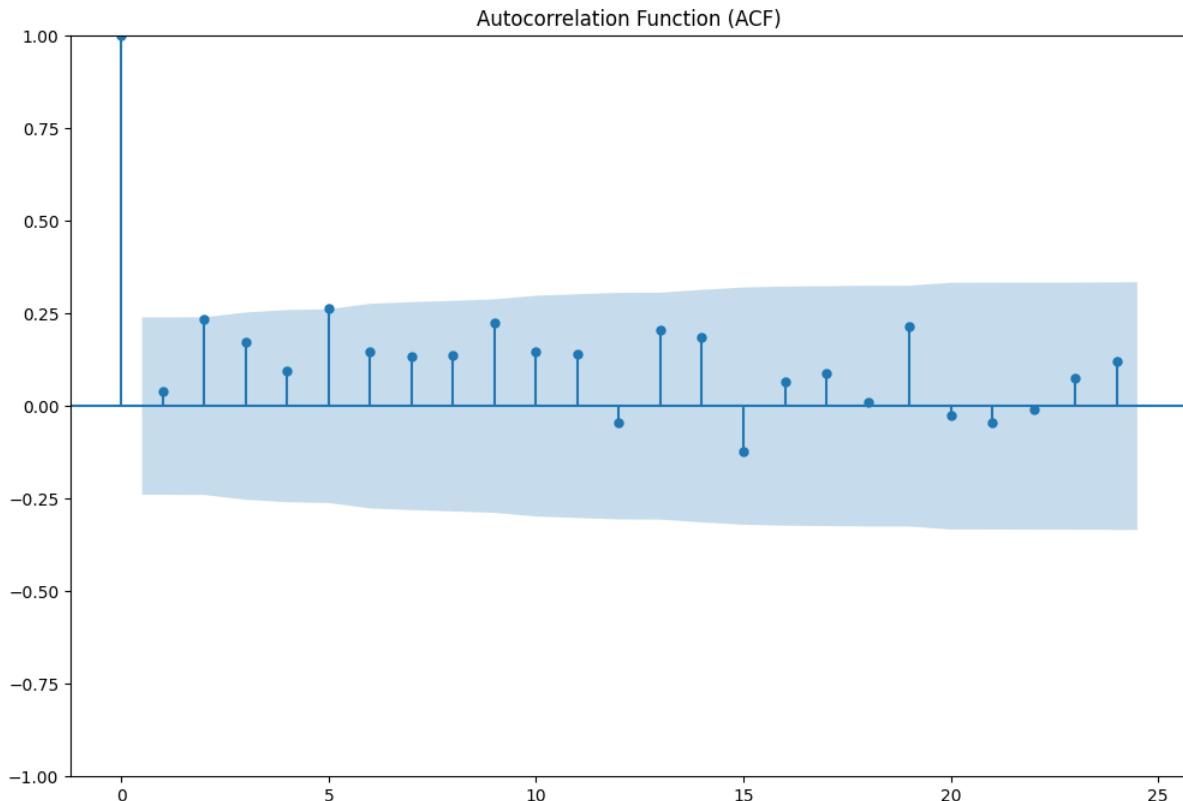
2.3. Componente residual

El componente residual oscila alrededor de cero, tal como se espera tras eliminar tendencia y estacionalidad. No obstante:

- Se identifican algunos valores atípicos (residuos de gran magnitud).
- Visualmente, el patrón residual no parece ser completamente aleatorio; es posible que persista cierta estructura temporal no capturada por los componentes anteriores.

Esto sugiere que, aunque la descomposición explica buena parte de la dinámica, podría existir aún correlación serial en los errores.

3. Análisis de autocorrelación



Para evaluar la dependencia temporal de la serie se emplean las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF).

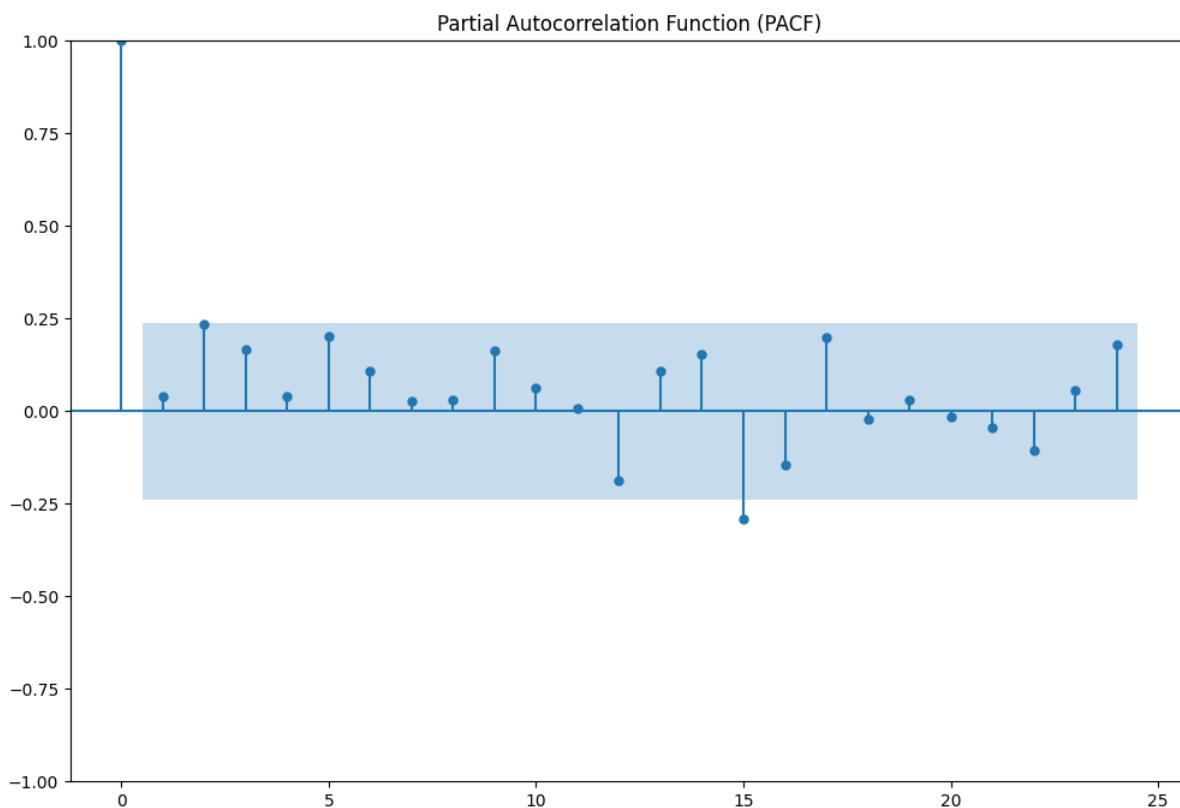
3.1. Función de Autocorrelación (ACF)

El correlograma ACF muestra:

- Valores de autocorrelación positivos y significativos en varios retardos (lags) iniciales.
- Picos relativamente destacados en algunos retardos concretos (por ejemplo, alrededor de los lags asociados al posible periodo estacional y sus múltiplos).

La lenta disminución y la presencia de picos en lags específicos indican que la serie presenta **memoria** y no puede considerarse ruido blanco. Además, la aparición de picos a intervalos regulares es coherente con la estructura estacional observada en la descomposición.

3.2. Función de Autocorrelación Parcial (PACF)



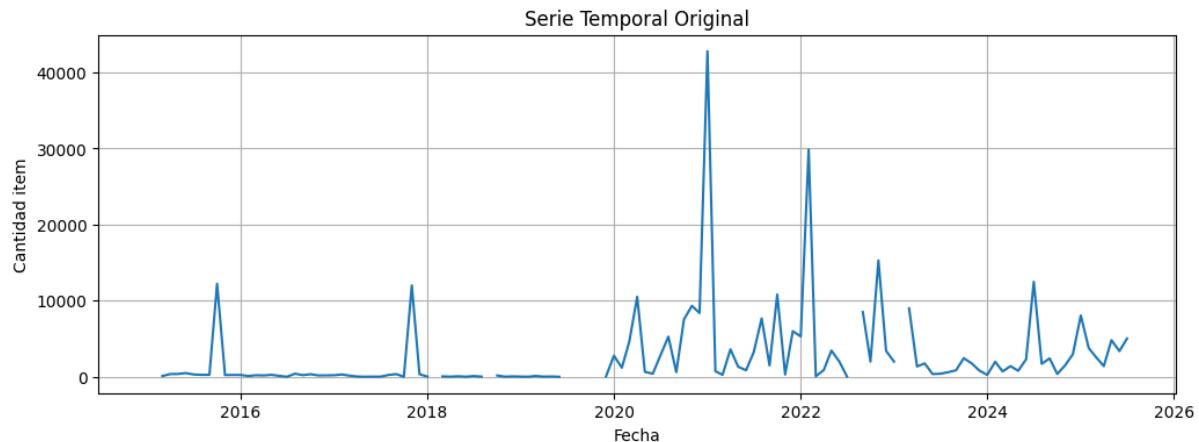
El correlograma PACF muestra:

- Varios lags iniciales (en particular, 1 y 2, y posiblemente algunos adicionales de orden bajo) con autocorrelaciones parciales significativas.
- Algunos picos adicionales en lags más alejados, compatibles con el periodo estacional de la serie.

Este comportamiento es típico de un proceso con un componente autorregresivo (AR) de orden bajo. La significancia de retardos asociados al periodo estacional sugiere, además, la conveniencia de incorporar términos AR estacionales en un modelo más completo.

Apósitos - El Salvador

Existen datos a partir del año 2025.



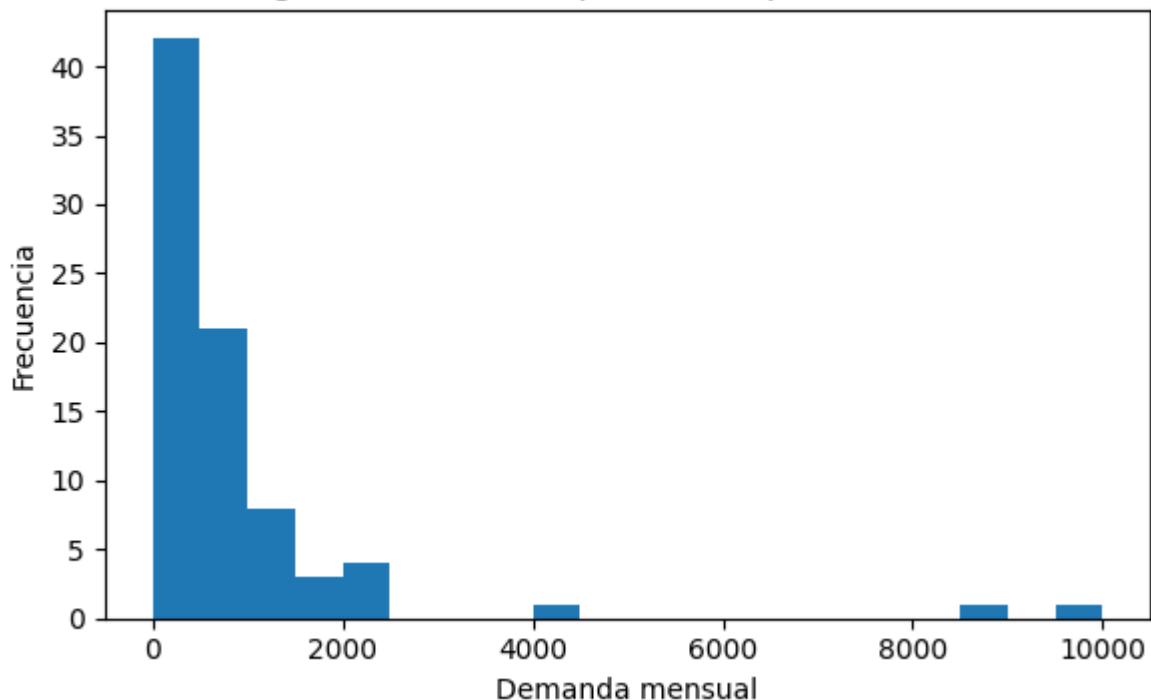
Existe mucho ruido en periodo de pandemia, se trató de experimentar utilizando ese periodo pero sin buenos resultados, por lo que se decide excluir esos datos y tomar órdenes a partir del año 2023.



- ✓ Primera señal de **demandas post-pandemia más “real” y estable**.
- ✓ Se aprecian variaciones, pero sin explosiones anómalas.
- ✓ Clave para forecasting: este es el período útil para predicción.

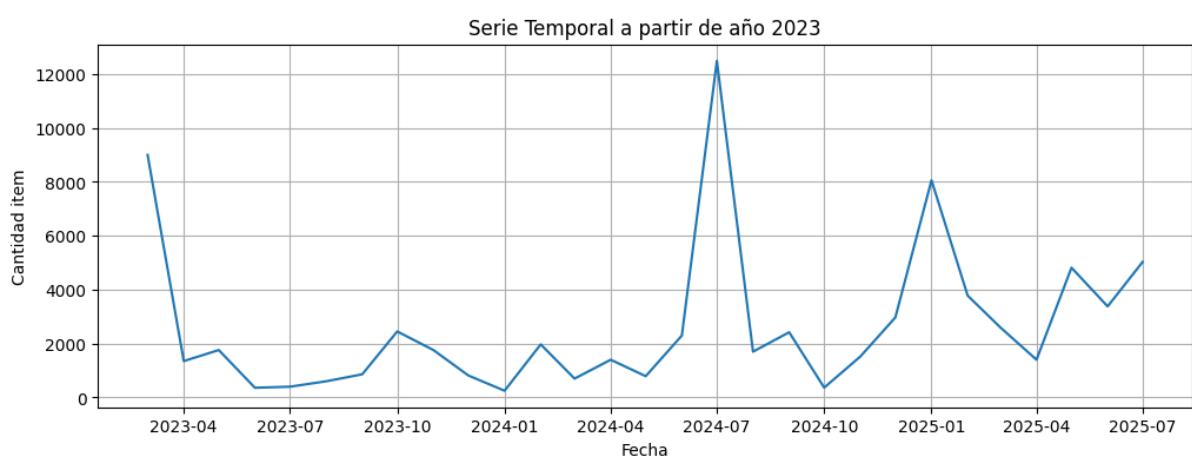
- ✓ No se excluyeron los outliers, ya que se perdía mucha información de los últimos períodos.

Histograma demanda apósitos hospital sótero del río



- ✓ Muestra nuevamente la asimetría, pero centrado en demanda mensual.
- ✓ Ayuda a justificar por qué se usan **modelos que trabajan con medianas o diferencias**.
- ✓ Valor agregado: permite explicar al negocio por qué ocurren meses “extremadamente altos”.

1. Descripción general de la serie

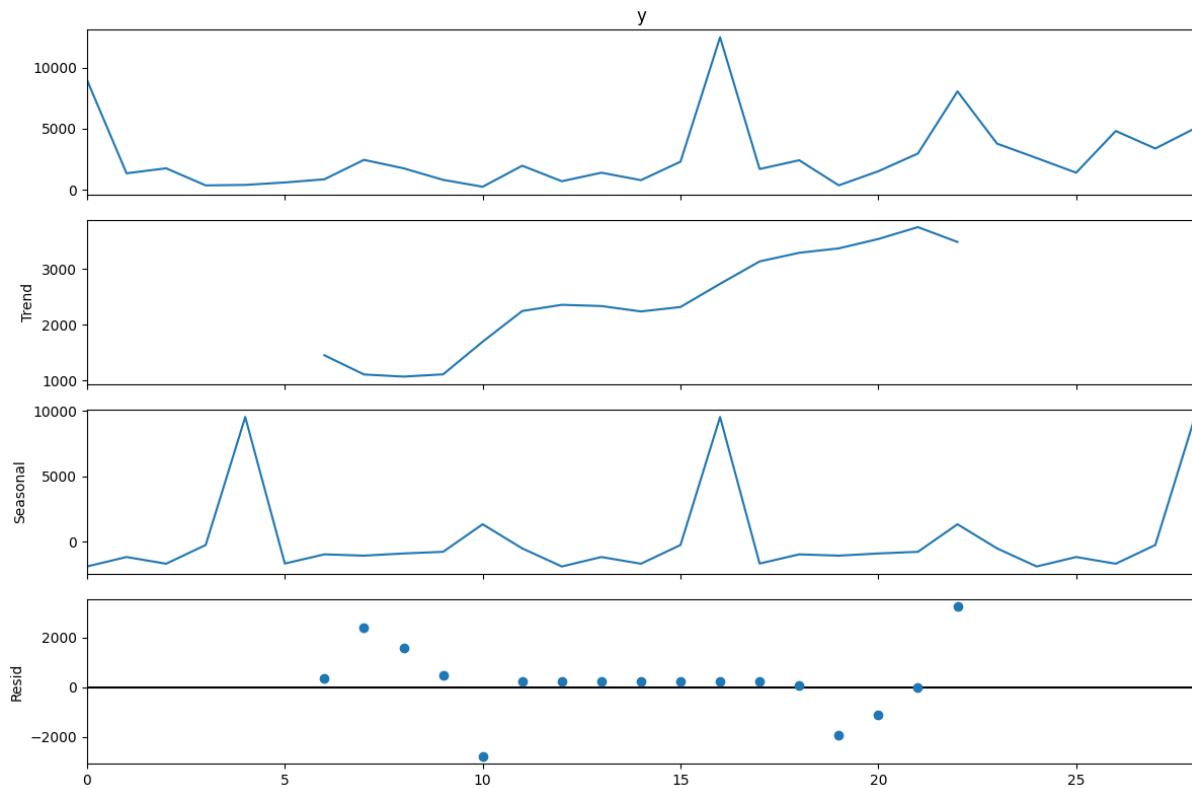


La serie observada presenta un número relativamente reducido de observaciones y se caracteriza por una alta variabilidad, con varios picos marcados de gran magnitud en puntos

específicos del horizonte temporal. Entre estos episodios de valores extremos, los niveles de la serie se mantienen en rangos moderados.

A simple vista no se aprecia un comportamiento puramente aleatorio: aunque existe ruido, se distinguen episodios de valores anómalamente altos que podrían interpretarse como eventos atípicos (outliers) o como manifestaciones de un componente estacional muy pronunciado en ciertas posiciones del ciclo.

2. Descomposición de la serie



La descomposición en componentes de tendencia, estacionalidad y residuo permite examinar la estructura interna del proceso.

2.1. Componente de tendencia

El componente de tendencia presenta un comportamiento **creciente** a lo largo del periodo en que está definido:

- En los primeros puntos su nivel se sitúa en torno a valores relativamente bajos.
- Posteriormente se observa un incremento progresivo hasta alcanzar niveles más elevados hacia el final del horizonte.

Este patrón indica que la serie no es estacionaria en nivel, dado que su media evoluciona con el tiempo. La presencia de una tendencia ascendente sugiere cambios estructurales en el proceso generador de datos (por ejemplo, crecimiento de la demanda, efecto acumulado de alguna política, etc.).

2.2. Componente estacional

El componente estacional muestra un patrón claramente recurrente y de gran amplitud:

- Aparecen picos positivos muy elevados en posiciones específicas, que se repiten aproximadamente con el mismo espaciamiento temporal.
- Entre estos picos, el componente estacional toma valores cercanos a cero o incluso negativos, generando un contraste marcado entre períodos “altos” y “bajos”.

Este comportamiento es consistente con una estacionalidad fuerte y concentrada, donde ciertos períodos registran valores sistemáticamente superiores al resto (por ejemplo, meses de alta demanda en un contexto anual). La amplitud de la estacionalidad es considerable en relación con el nivel medio de la serie.

2.3. Componente residual

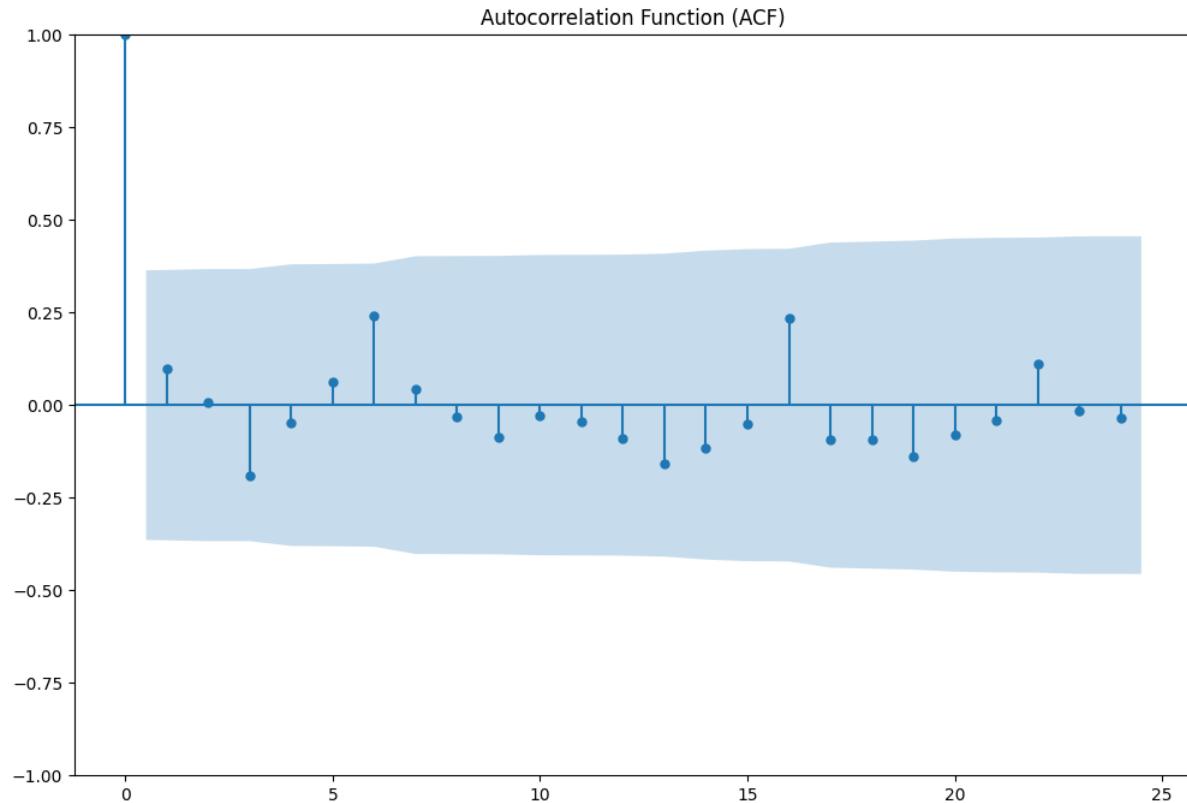
El componente residual oscila alrededor de cero, pero no se comporta como ruido blanco puro:

- En la parte inicial se observan principalmente residuos positivos de magnitud moderada.
- En la zona intermedia y final aparecen residuos predominantemente negativos, algunos de ellos de gran magnitud.

Esta posible asimetría temporal en los residuos sugiere que aún persiste cierta estructura no capturada por la combinación de tendencia y estacionalidad, o bien que existen episodios puntuales que podrían modelarse como outliers o cambios de nivel.

3. Análisis de autocorrelación

3.1. Función de Autocorrelación (ACF)

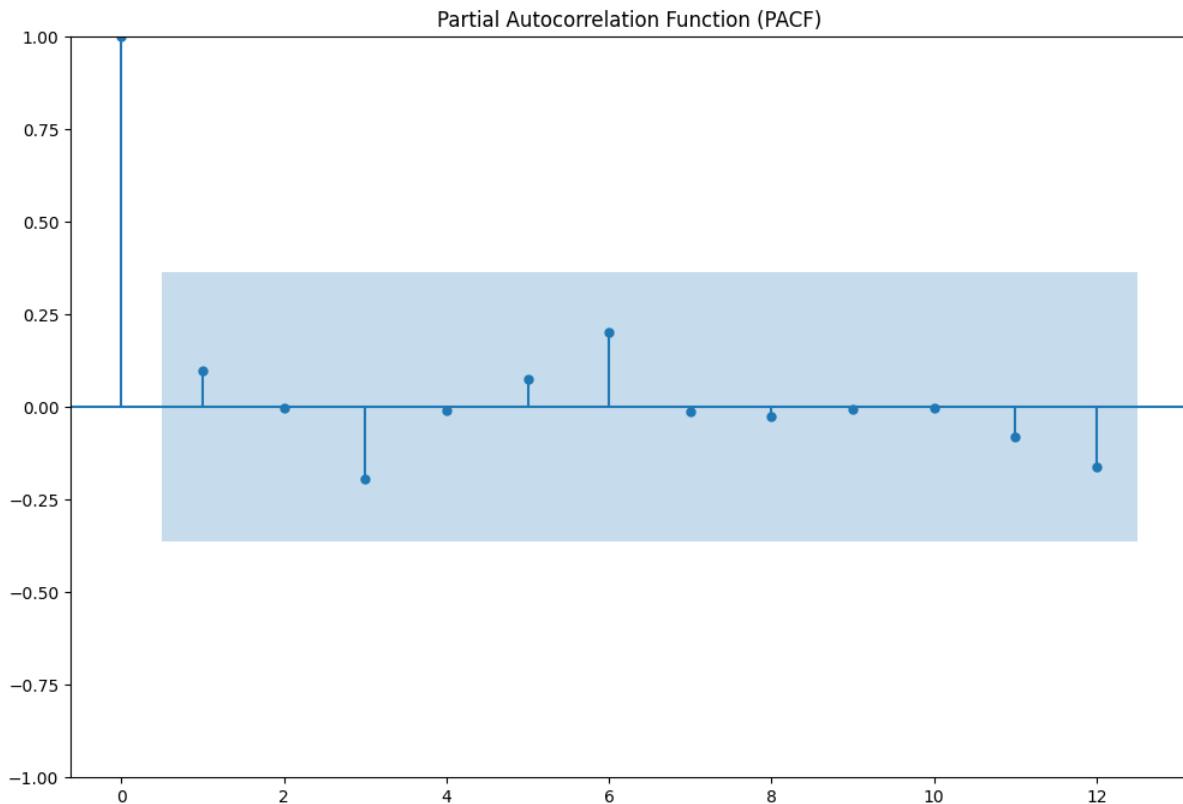


El corregograma ACF de la serie muestra:

- La mayoría de las autocorrelaciones para lags mayores que 0 se sitúan dentro de las bandas de confianza, es decir, no son estadísticamente significativas.
- Se observan algunos coeficientes aislados de mayor magnitud (por ejemplo, en ciertos lags bajos y en lags asociados al posible periodo estacional), pero no existe un patrón de decrecimiento lento ni una estructura clara de tipo AR o MA.

En conjunto, la ACF sugiere que la serie exhibe dependencia temporal limitada: no hay evidencia de una fuerte memoria de corto plazo, aunque se aprecian indicios puntuales de estacionalidad en algunos retardos específicos.

3.2. Función de Autocorrelación Parcial (PACF)



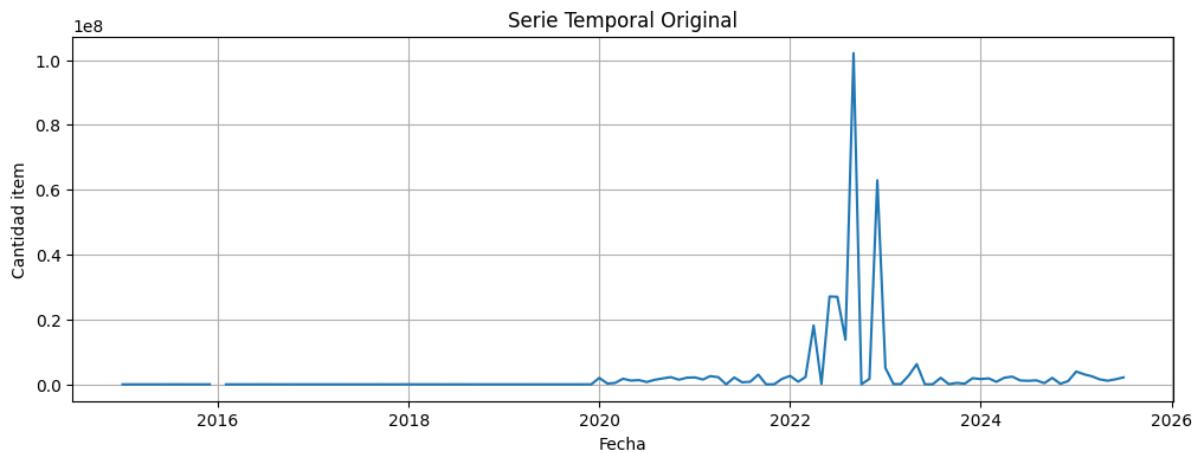
El correlograma PACF muestra:

- Coeficientes parciales mayormente contenidos dentro de las bandas de confianza.
- No se aprecia un corte claro en los primeros lags ni un comportamiento típico de un proceso autorregresivo de orden bajo (por ejemplo, AR(1) o AR(2)) fuertemente dominante.
- Algunos coeficientes negativos en lags intermedios podrían indicar ligeras correcciones de corto plazo, pero su magnitud es moderada.

Este patrón es compatible con una serie donde el componente autorregresivo es débil o de orden muy bajo, y donde gran parte de la dinámica es explicada por la tendencia, la estacionalidad y la presencia de valores extremos.

Guantes - Sótero del Río

La serie tiene datos desde el año 2015.

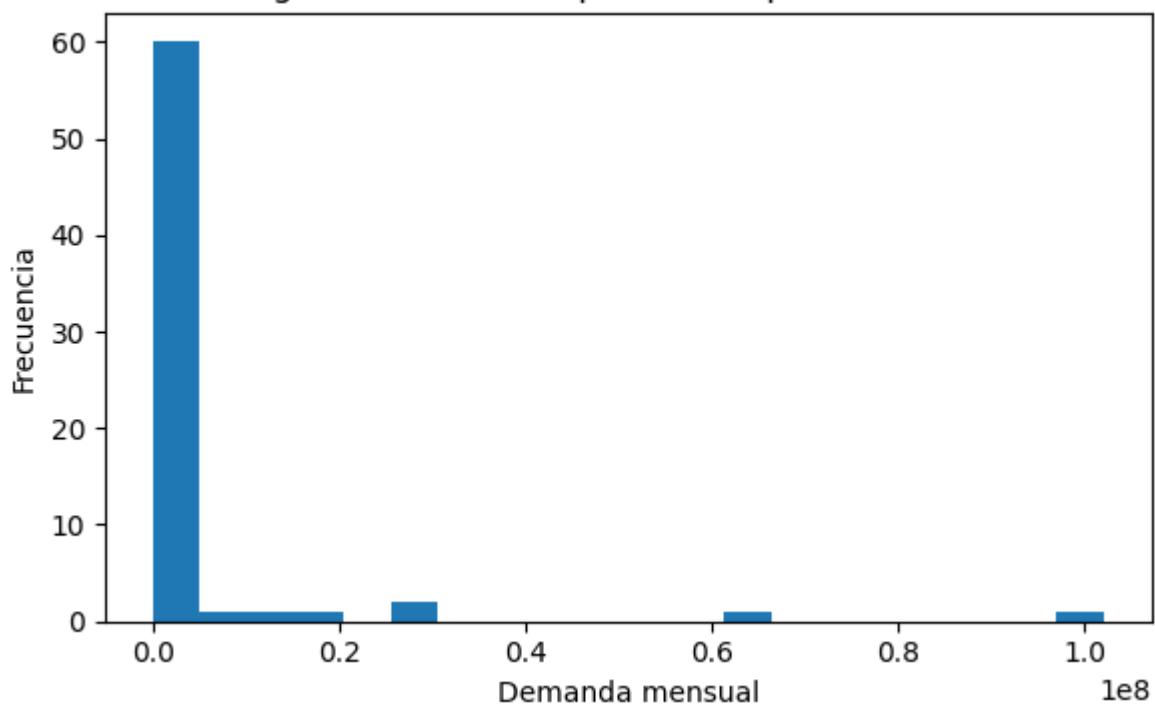


Se filtran los datos anteriores a 2020.

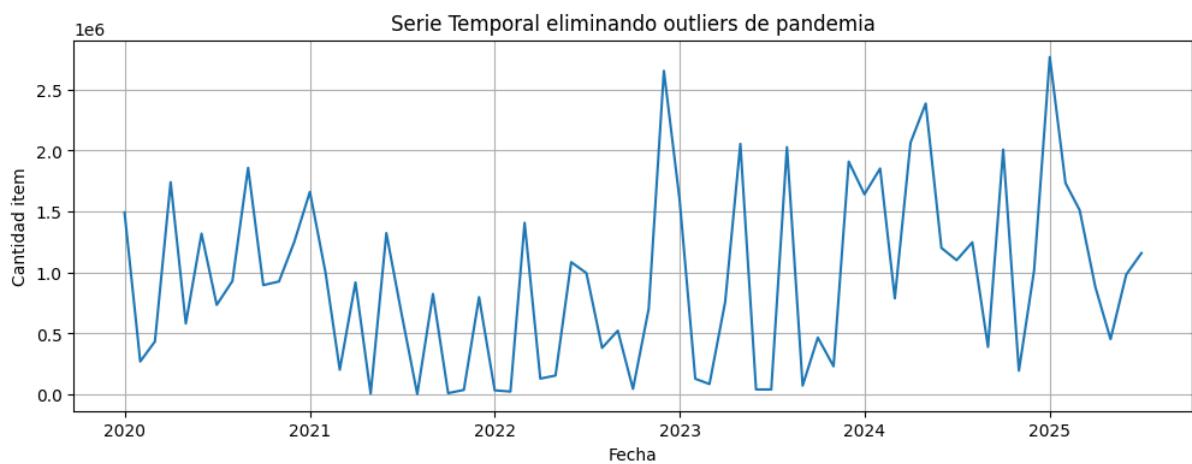


Existe un comportamiento distinto en el periodo de pandemia, por lo que se van a eliminar los valores atípicos para poder utilizar este periodo.

Histograma demanda apósitos hospital sótero del río



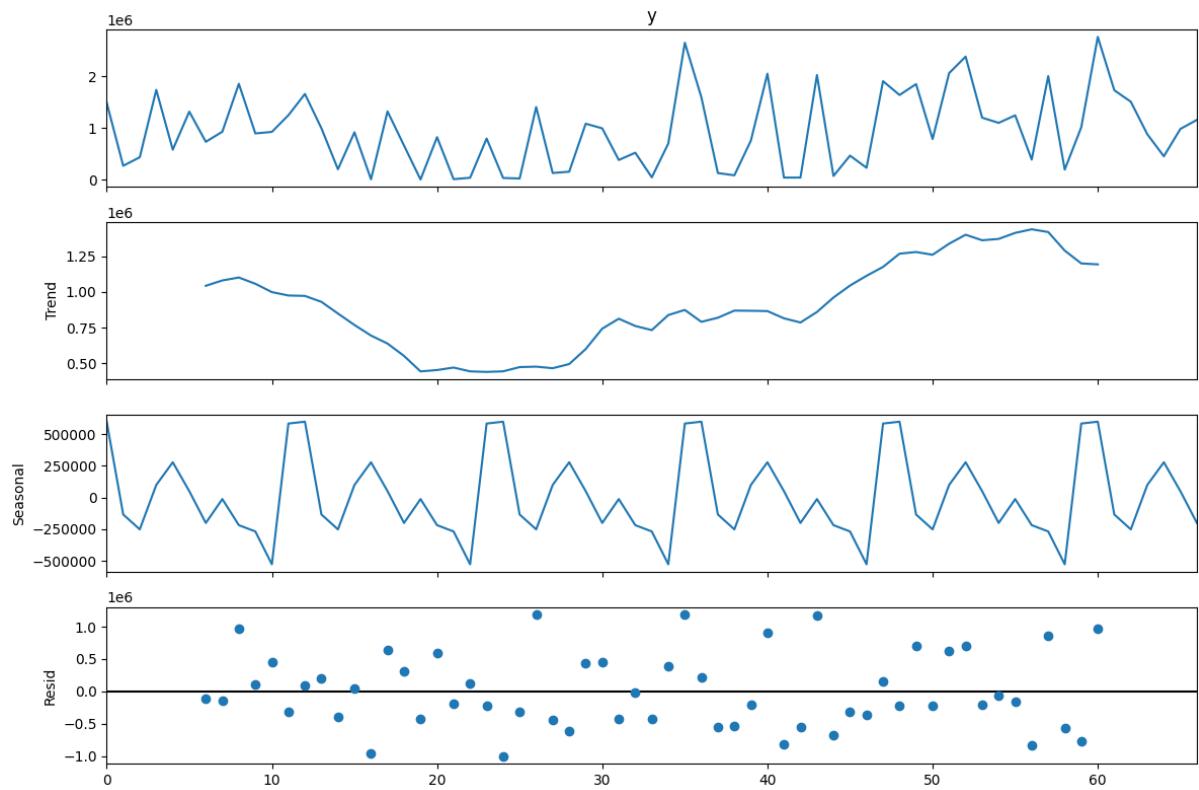
1. Descripción general de la serie



La serie original presenta valores del orden de millones de unidades, con oscilaciones marcadas a lo largo de todo el horizonte temporal. Se observan incrementos y descensos relativamente bruscos, así como episodios de valores máximos localizados.

A simple vista, el proceso no parece ser puramente aleatorio: las fluctuaciones muestran cierta regularidad y la media de la serie parece cambiar en distintos tramos, lo que sugiere la presencia conjunta de tendencia y estacionalidad.

2. Descomposición de la serie



La descomposición de la serie en componentes de tendencia, estacionalidad y residuo permite examinar de manera más precisa la estructura subyacente.

2.1. Componente de tendencia

El componente de tendencia exhibe un comportamiento no lineal a lo largo del tiempo:

- En los primeros períodos la tendencia se sitúa en torno a un nivel ligeramente superior a 1×10^6 .
- Posteriormente se observa un descenso progresivo hasta alcanzar un mínimo cercano a 0.45×10^6 .
- A partir de ese punto, la tendencia se recupera y muestra una trayectoria creciente, superando nuevamente el millón de unidades y alcanzando niveles máximos hacia el final del horizonte, con una ligera estabilización o corrección a la baja en los últimos puntos.

Esta evolución indica que la serie es no estacionaria en nivel, dado que su media cambia de forma sistemática en el tiempo.

2.2. Componente estacional

El componente estacional presenta un patrón claramente periódico y de gran amplitud:

- Se observan ciclos que se repiten de manera regular, con picos positivos y negativos de magnitud considerable (del orden de $\pm 0.5 \times 10^6$).
- El patrón se repite de forma bastante estable a lo largo de la serie, lo que evidencia una estacionalidad fuerte asociada a un periodo fijo (por ejemplo, mensual o trimestral, dependiendo de la frecuencia original de los datos).

La magnitud de este componente es relevante en relación con el nivel medio de la serie, por lo que la estacionalidad constituye un elemento central en la dinámica del proceso.

2.3. Componente residual

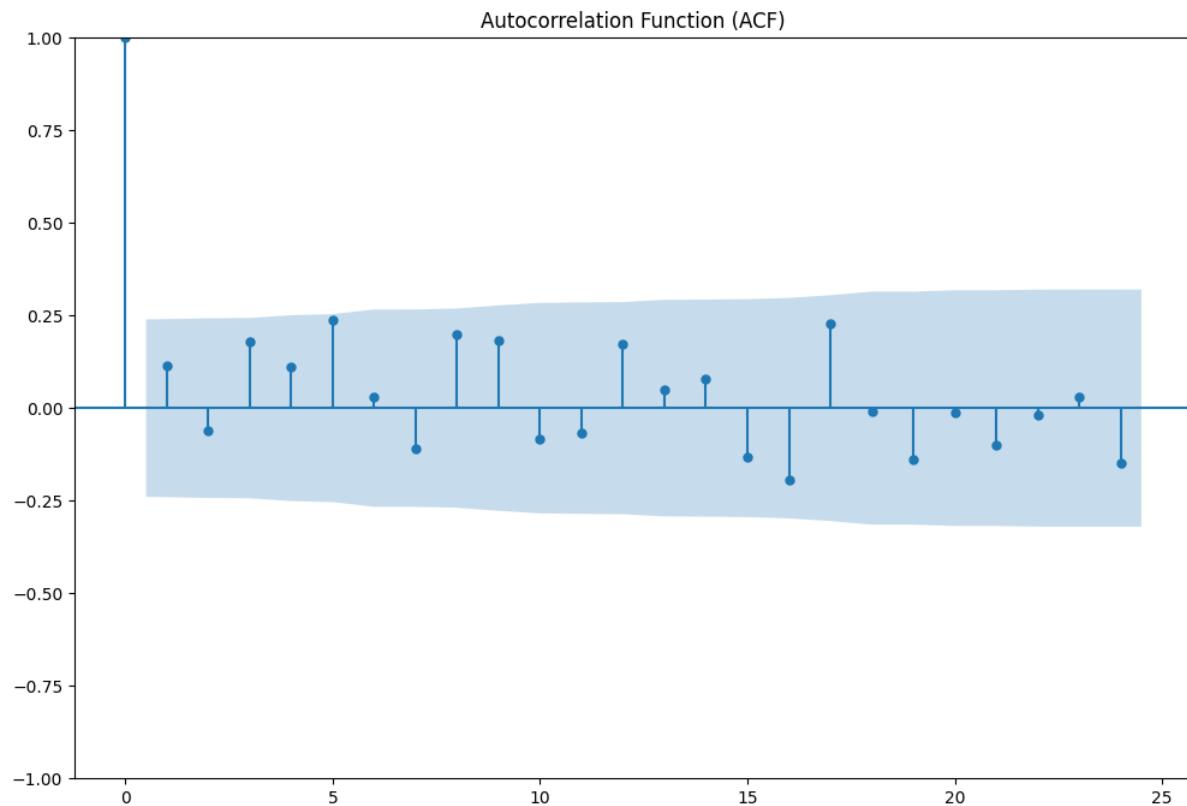
El componente residual oscila alrededor de cero y muestra una dispersión considerable:

- Los residuos se distribuyen tanto en valores positivos como negativos, con algunos puntos de gran magnitud.
- Visualmente no se aprecian patrones deterministas evidentes (como tendencia residual marcada o estacionalidad remanente), aunque la presencia de valores extremos sugiere posibles **outliers** o episodios puntuales no explicados por los componentes de tendencia y estacionalidad.

En conjunto, los residuos se aproximan a un comportamiento de ruido, si bien es necesario contrastar formalmente la ausencia de autocorrelación mediante pruebas específicas.

3. Análisis de autocorrelación

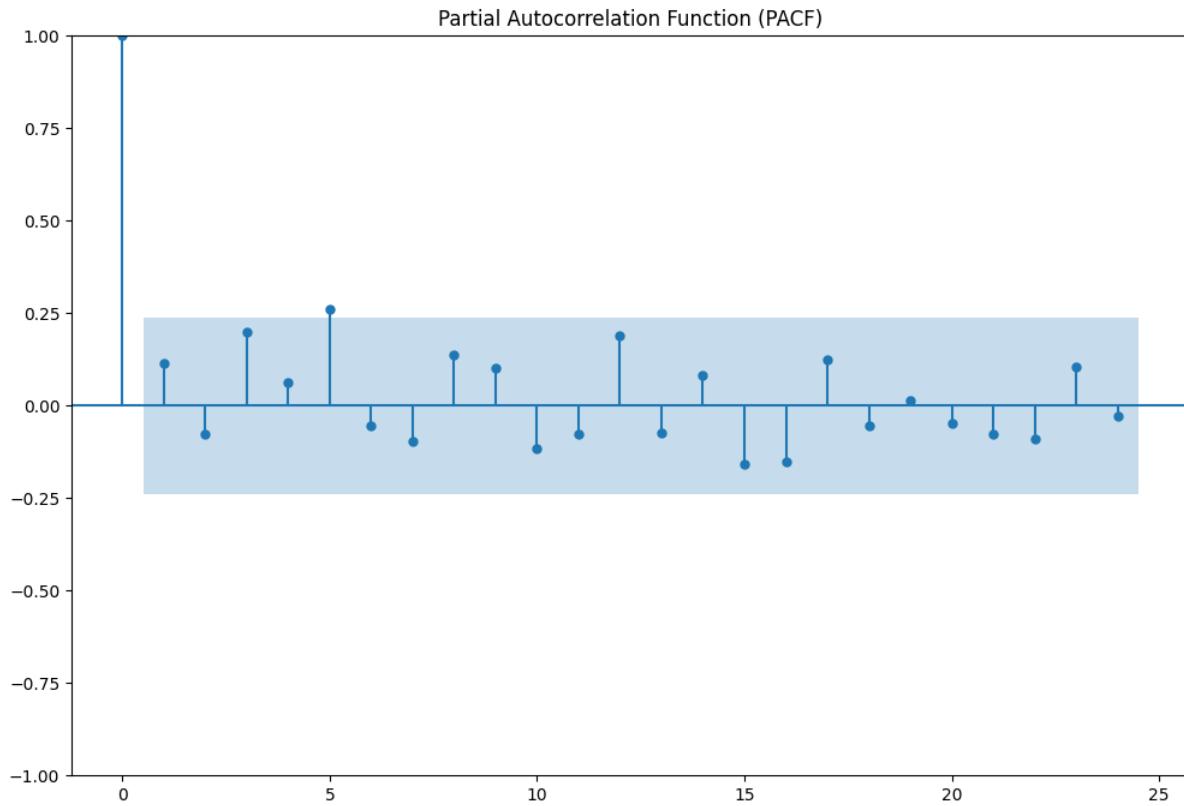
3.1. Función de Autocorrelación (ACF)



El corregograma ACF refleja el grado de dependencia serial de la serie:

- En los primeros lags se observan autocorrelaciones positivas y de magnitud moderada, algunas de ellas cercanas al límite de las bandas de confianza, lo que indica cierta persistencia de corto plazo.
- A partir de lags intermedios aparecen valores tanto positivos como negativos, en general de menor magnitud; algunos lags aislados superan ligeramente las bandas de confianza, lo que podría vincularse con la estacionalidad.
- No se observa un decaimiento lento y monótono típico de procesos con fuerte memoria de largo plazo, pero la existencia de varios lags significativos confirma que la serie no es ruido blanco.

3.2. Función de Autocorrelación Parcial (PACF)



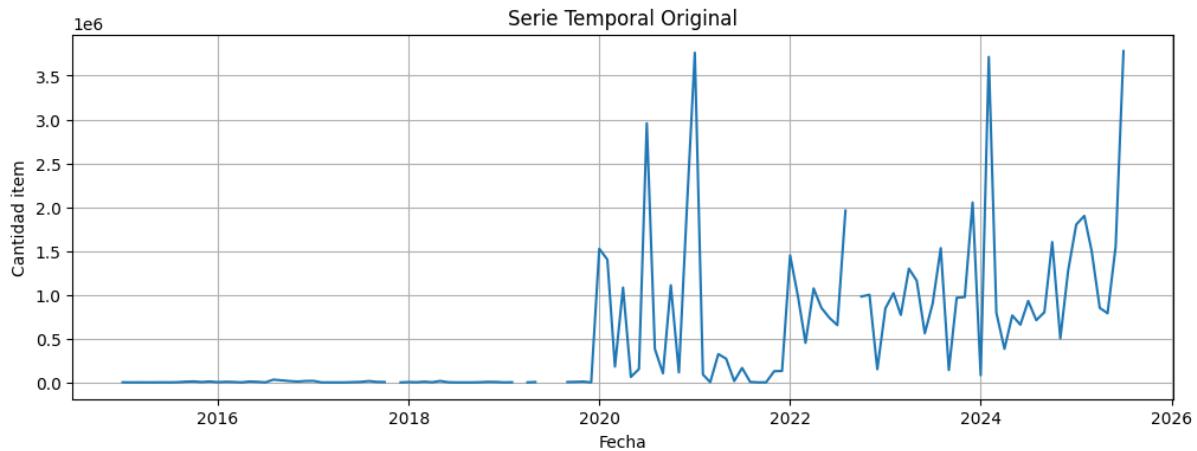
El correlograma PACF proporciona información sobre las dependencias directas entre la serie y sus retardos:

- Los primeros lags (en particular, los 1–3 y algún lag adicional de orden bajo) presentan autocorrelaciones parciales positivas y cercanas o ligeramente superiores a las bandas de confianza, compatibles con la presencia de un componente autorregresivo de orden bajo.
- En lags más elevados, los coeficientes se mantienen mayoritariamente dentro de las bandas, con algunos valores negativos de pequeña a moderada magnitud, lo que indica que la contribución de retardos lejanos es limitada.

La PACF sugiere que un modelo autorregresivo de orden reducido podría capturar buena parte de la dependencia de corto plazo, complementando el tratamiento de la tendencia y la estacionalidad.

Guantes - El Salvador

Existen datos desde el 2015, con muy bajo volumen, recién en el año 2020 comienza a existir un mayor volumen.

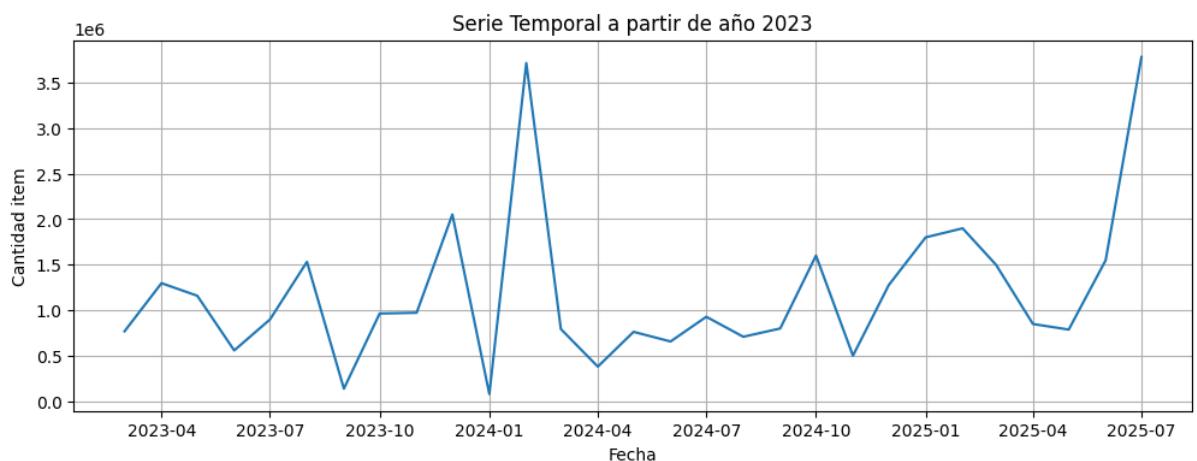


Dentro del periodo de pandemia existen patrones muy aleatorios que no se pudieron excluir, por lo que se decidió sacar ese periodo.

✓ Relevancia comercial:

- Confirma que múltiples productos presentan patrones similares → La pandemia impulsó compras masivas.

✓ Apoyo al argumento: **la demanda actual ya no se parece a la histórica pre-pandemia.**



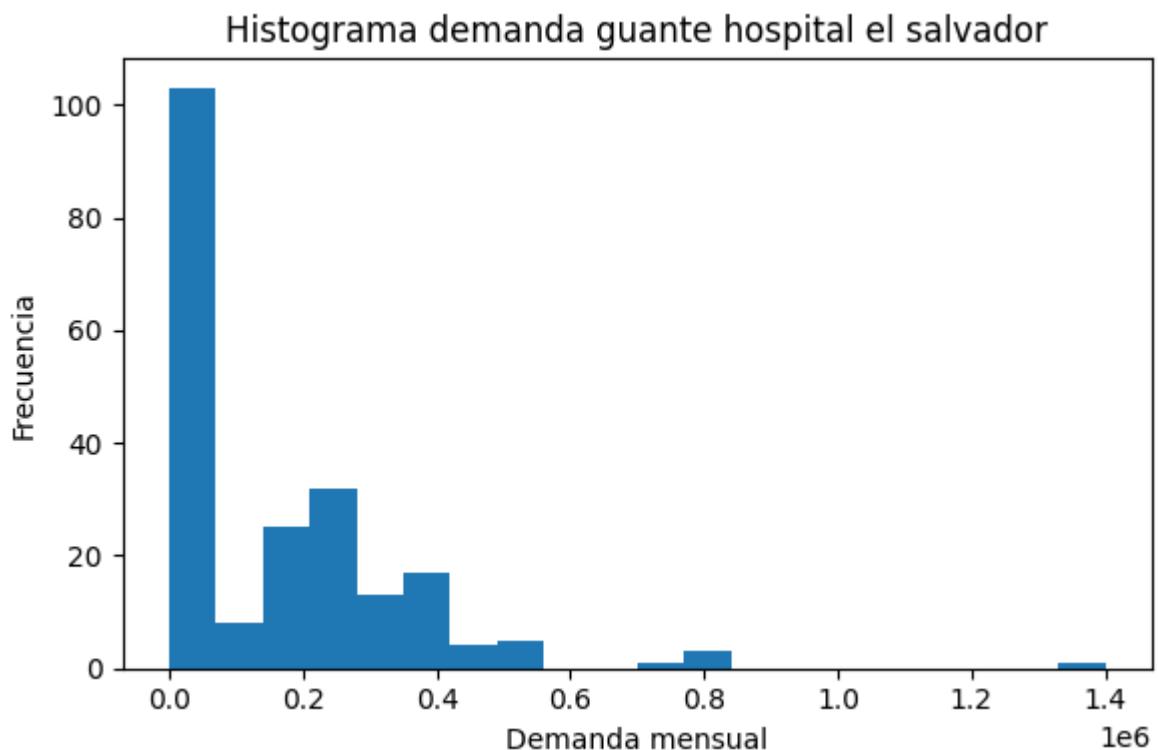
✓ Observamos una demanda mucho mayor que apóstitos y más variable.

✓ Strategic Insight:

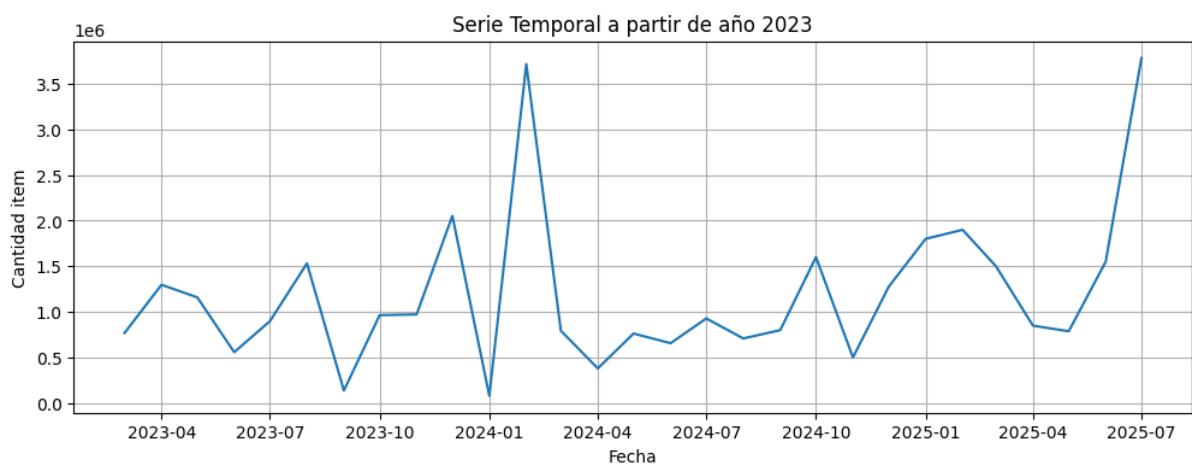
- Este producto parece crítico y de alto volumen.
- Mayor volatilidad → riesgo de quiebres de stock si no se pronostica bien.

✓ Conexión económica:

- Potencial de ahorro alto con forecast certero.

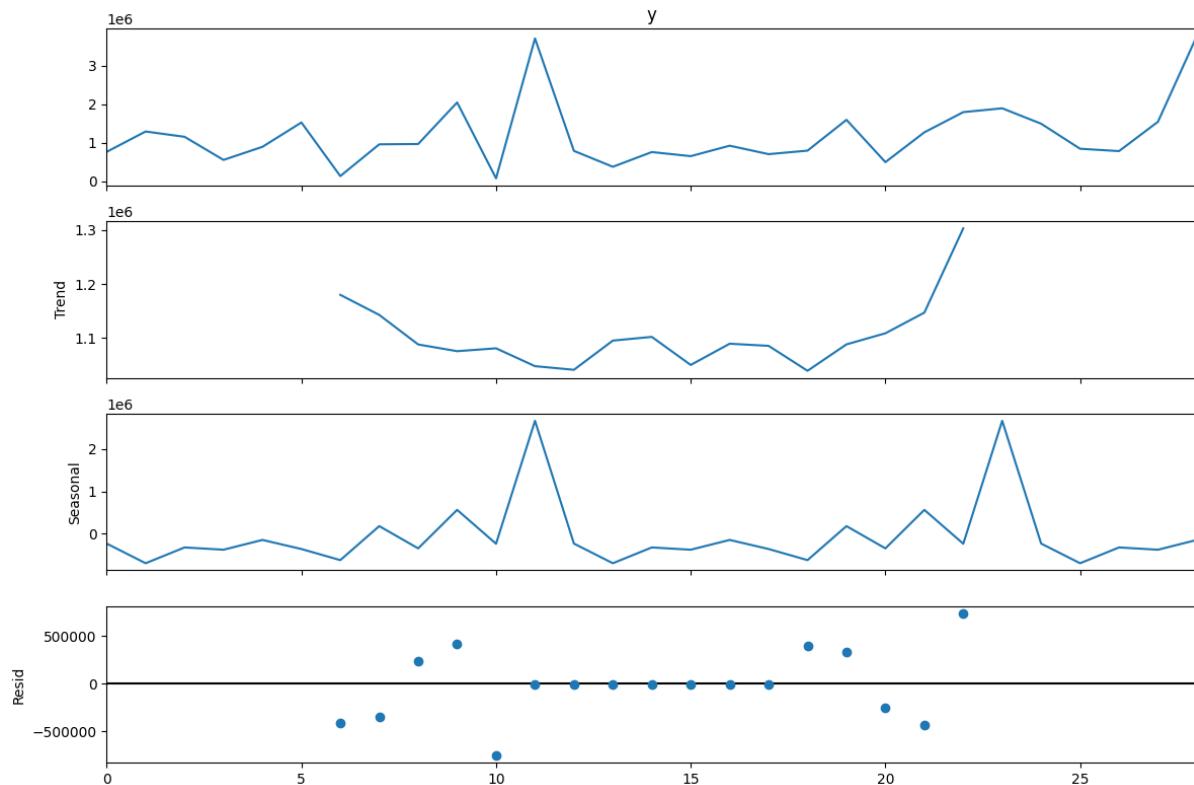


1. Descripción general de la serie



La serie original presenta valores del orden de millones unidades, con oscilaciones de magnitud moderada a lo largo del horizonte temporal. Se observan varios picos y valles, pero el comportamiento general sugiere que el proceso no es puramente aleatorio: la media de la serie parece incrementarse levemente en los últimos períodos, lo cual es indicio de no estacionariedad en nivel.

2. Descomposición de la serie



La descomposición en componentes de tendencia, estacionalidad y residuo permite caracterizar de forma más precisa la dinámica de la serie.

2.1. Componente de tendencia

El componente de tendencia muestra un comportamiento básicamente creciente, aunque con pequeñas fluctuaciones:

- En los primeros períodos, la tendencia se sitúa en torno a 1×10^6 .
- Hacia la mitad de la serie se observa una ligera disminución, seguida de una recuperación progresiva.
- En los tramos finales, la tendencia asciende de manera sostenida hasta ubicarse por encima de 1.1×10^6 .

Este patrón confirma que la serie presenta cambios sistemáticos en su nivel medio, y por tanto no es estacionaria.

2.2. Componente estacional

El componente estacional presenta variaciones de amplitud moderada:

- Se observan picos positivos y negativos, pero el patrón no es tan regular ni tan marcado como en otros casos; es decir, la estacionalidad existe, pero su intensidad es relativamente limitada.
- La forma de los ciclos sugiere cierta repetición, aunque con mayor irregularidad en magnitud y duración que en series fuertemente estacionales.

En consecuencia, la estacionalidad parece desempeñar un papel complementario, pero no dominante, en la explicación de la serie.

2.3. Componente residual

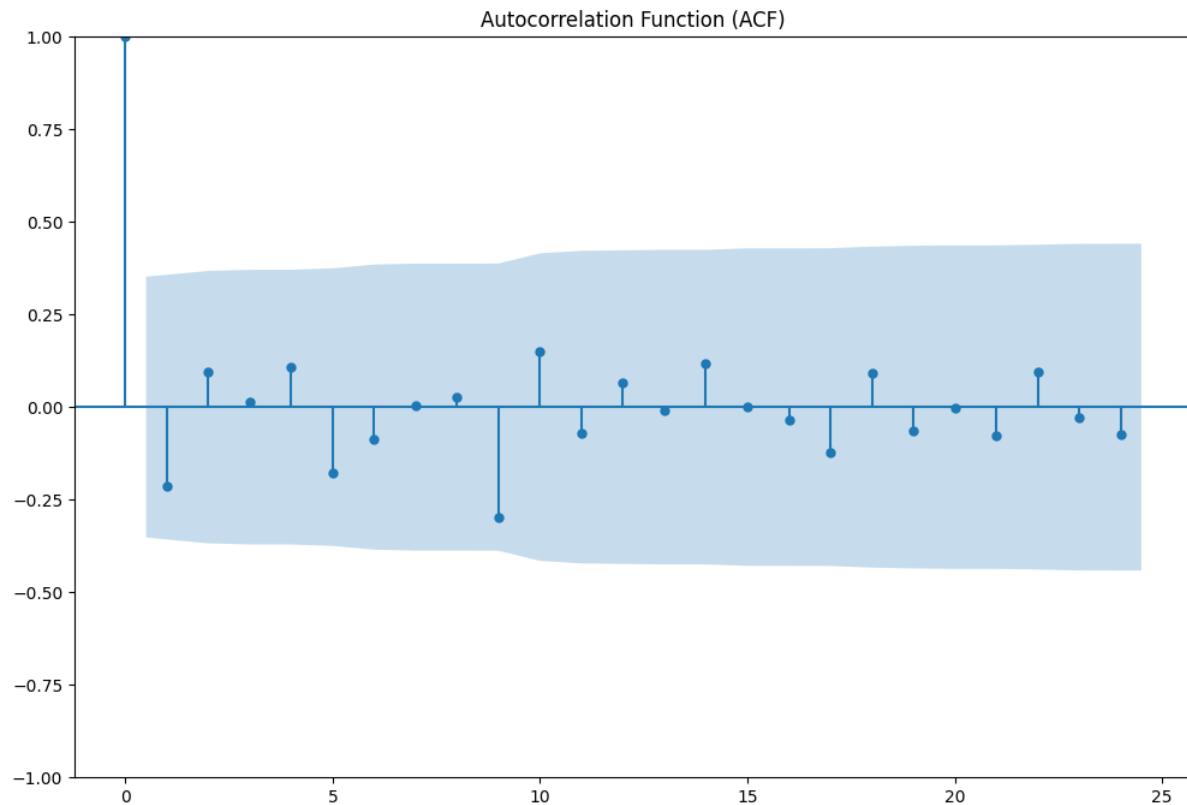
El componente residual oscila alrededor de cero y presenta:

- Algunos valores positivos y negativos de gran magnitud en puntos concretos, que podrían interpretarse como observaciones atípicas o perturbaciones de gran tamaño.
- Una porción central de la serie en la que los residuos se concentran cerca de cero, lo que indica que, en ese tramo, la combinación de tendencia y estacionalidad captura razonablemente bien el comportamiento observado.

En términos generales, los residuos no muestran un patrón determinista evidente, aunque la presencia de outliers aconseja un análisis adicional (por ejemplo, mediante modelos de intervención).

3. Análisis de autocorrelación

3.1. Función de Autocorrelación (ACF)

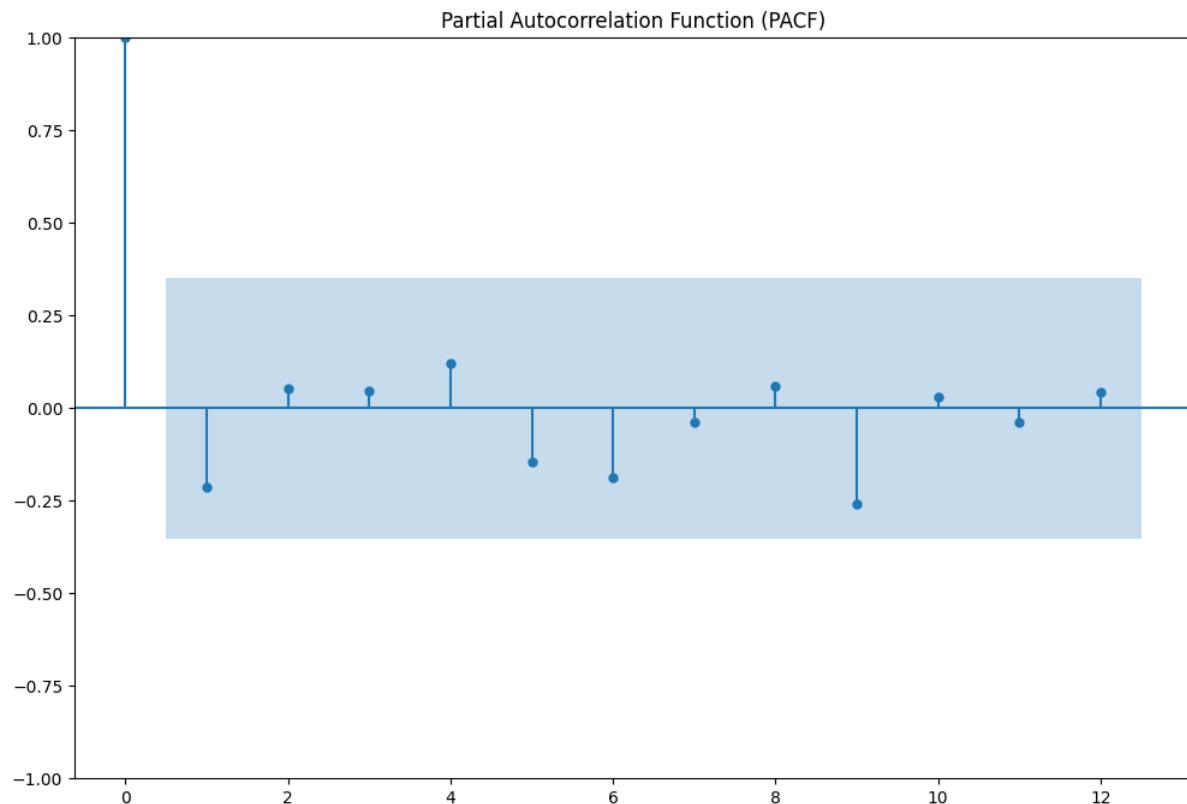


El corregograma ACF presenta las siguientes características:

- El coeficiente de autocorrelación en el lag 1 es claramente negativo y se sitúa fuera de las bandas de confianza, lo que indica una relación inversa de corto plazo entre observaciones consecutivas.
- El resto de los lags muestran autocorrelaciones de magnitud baja o moderada, mayoritariamente dentro de las bandas de confianza, con alguna excepción puntual (por ejemplo, un valor negativo algo más pronunciado en torno al lag 9).
- No se observa un decaimiento lento ni un patrón repetitivo fuerte, lo que sugiere una dependencia serial limitada.

En conjunto, la ACF es compatible con un proceso con memoria de corto alcance, posiblemente con un componente de tipo MA de bajo orden.

3.2. Función de Autocorrelación Parcial (PACF)



El correlograma PACF muestra:

- Un lag 1 claramente significativo y negativo, que refuerza la idea de una relación directa (en sentido inverso) entre la observación actual y la inmediatamente anterior.
- Para lags superiores, los coeficientes parciales se mantienen dentro de las bandas de confianza y sus magnitudes son pequeñas, lo que indica que la contribución de retardos de orden mayor es reducida.

Este patrón es típico de procesos en los que la dinámica puede describirse adecuadamente mediante un modelo MA(1) (o, en el contexto de modelos ARIMA, un componente MA de orden 1).

Etapa 4: Agregación y generación de dataset de entrenamiento

Una vez finalizado el tratamiento de las series de tiempo en la fase de análisis exploratorio — donde se seleccionaron las ventanas temporales de mayor relevancia comercial para la etapa post-pandemia y se gestionaron los valores atípicos (outliers) que distorsionaban la señal real de consumo— se procede a la construcción de las variables predictoras (features) necesarias para habilitar el aprendizaje de los modelos de Machine Learning.

El objetivo de esta etapa es transformar los datos transaccionales depurados en una estructura consolidada que permita realizar un análisis robusto y consistente entre los distintos hospitales.

Pasos detallados:

a) Agrupación de registros históricos:

- Se agrupan los datos por combinación de:
 - **Institucion** (hospital comprador)
 - **producto_id** (producto estandarizado)
 - **año_mes** (mes de referencia)
- Esta agrupación permite obtener un resumen mensual de la demanda por hospital Producto.

b) Generación de la variable objetivo (demandas_mensual**):**

- Se calcula la **suma de CantidadItem** en cada grupo, lo que representa la demanda total mensual del producto por institución.

c) Construcción de variables predictoras (features):

Para enriquecer el dataset y entregar a los modelos información relevante que facilite anticipar la demanda futura, se generó un conjunto de variables derivadas a partir de rezagos, transformaciones estadísticas y características temporales. La implementación de estas transformaciones se realizó empleando el ecosistema Nixtla,

específicamente los módulos de ingeniería de atributos provistos por *StatsForecast* y *MLForecast*, debido a su soporte para procesamiento estrictamente causal y reproducible en series temporales (Garza et al., 2022).

1. Variables basadas en rezagos.

Se incorporaron rezagos directos de la demanda para capturar dependencia temporal:

- **lag_1, lag_2, lag_3, lag_6:** valores de la demanda de 1, 2, 3 y 6 meses atrás.

Adicionalmente, se aplicaron transformaciones sobre estos rezagos para capturar tendencias locales y variabilidad reciente:

- **Promedios móviles:**
 - Promedio móvil de 3 meses aplicado a *lag 1* y *lag 2*, para suavizar fluctuaciones de corto plazo.
- **Desviación estándar móvil:**
 - Desviación estándar de 3 meses sobre *lag 3*, permitiendo capturar la variabilidad reciente en la demanda.
- **Combinación de rezagos:**
 - Diferencia entre *lag_1* y *lag_2*, útil para representar cambios inmediatos en el nivel de demanda.

Todas estas variables fueron generadas con las funciones de *feature engineering* de Nixtla, que garantizan que cada cálculo utilice únicamente información disponible antes del período objetivo ($t-1, t-2, \dots$), evitando cualquier fuga de información (*data leakage*) y asegurando un proceso estrictamente causal.

2. Transformación del target

- **Diferenciación de orden 1:**
 Se aplicó una diferencia simple de orden 1 para estabilizar la serie temporal y reducir su autocorrelación, lo que mejora la capacidad predictiva de los modelos supervisados implementados en *MLForecast*. Esta transformación es gestionada de manera interna por el módulo de Nixtla mediante *Differences()*, el cual también realiza automáticamente la inversión de la transformación al

momento de generar las predicciones. Esto se logra sumando las predicciones diferenciadas al valor observado del período previo, permitiendo recuperar la escala original de la demanda. Gracias a este mecanismo integrado, las métricas y los pronósticos finales se expresan directamente en unidades reales, sin necesidad de procedimientos adicionales de post-procesamiento.

3. Variables temporales

Se incorporaron *date features* automáticamente a partir de la fecha del registro:

- **month**: mes del año (1–12)
- **quarter**: trimestre (1–4)
- **year**: año calendario

Estas variables permiten a los modelos capturar estacionalidades y patrones recurrentes asociados al calendario.

Resultado:

Se obtiene una tabla consolidada con una fila por cada combinación (*hospital*, *producto*, *mes*), que incluye:

- La demanda mensual en su escala original.
- Variables basadas en rezagos y sus transformaciones, generadas de forma causal mediante Nixtla.
- Variables temporales que capturan estacionalidades.

Este diseño proporciona a los modelos una representación temporalmente informada y metodológicamente sólida, apoyada en herramientas especializadas para forecasting que garantizan consistencia, reproducibilidad y ausencia de *data leakage*.

Etapa 5: Modelado

Previo al proceso de entrenamiento, se definió la estrategia de modelamiento considerando la naturaleza heterogénea de las series. En este estudio se adoptó un enfoque basado en modelos individuales por combinación hospital–producto, fundamentado en la marcada variabilidad estructural observada entre hospitales, tanto en volúmenes de demanda como en protocolos clínicos, prácticas de abastecimiento y dinámicas estacionales.

Esta heterogeneidad implica que las series no comparten patrones temporales consistentes que permitan el uso eficiente de un modelo global multiserie. En consecuencia, la estimación separada de cada serie facilita la captura de sus características locales —incluyendo tendencias específicas, shocks idiosincráticos y cambios estructurales dependientes del contexto organizacional de cada hospital— favoreciendo una mejor capacidad de ajuste y una interpretación más clara de los resultados.

La literatura de pronósticos respalda este enfoque en contextos donde las series presentan alta diversidad y escasa transferibilidad de patrones. Hyndman y Athanasopoulos (2021) argumentan que, cuando las series individuales difieren sustancialmente en sus dinámicas temporales o en los factores que las generan, los modelos independientes suelen ofrecer un desempeño superior en comparación con los modelos globales, dado que permiten una representación más fiel de la estructura temporal inherente a cada unidad de análisis.

Además, se estableció un horizonte de predicción de 3 meses, definido en función de las necesidades operativas del equipo comercial. Este período corresponde al tiempo mínimo requerido para planificar campañas, coordinar intervenciones en los hospitales y gestionar los procesos logísticos asociados al abastecimiento. Por lo tanto, el horizonte seleccionado no solo es metodológicamente razonable, sino también alineado con la ventana estratégica necesaria para una toma de decisiones informada y oportuna.

Una vez construido el dataset consolidado, se procede al entrenamiento de modelos de regresión con el objetivo de predecir la demanda mensual esperada por combinación hospital–producto.

- Se evaluaron **modelos estadísticos clásicos y modelos de Machine Learning**.
- Selección basada en:
 - Su uso frecuente en la literatura.

- **Efectividad comprobada** en aplicaciones reales.
- Rol como **baselines** en competencias internacionales (M4, M5).

Modelos Estadísticos: Usados por su capacidad para **capturar patrones estructurales**: tendencia, estacionalidad y ciclos.

- **ARIMA**
 - Combina componentes autorregresivos (AR), de media móvil (MA) y diferenciación (I).
 - Modela tendencias y ciclos en la serie temporal.
- **CES (Complex Exponential Smoothing)**
 - Variante del suavizamiento exponencial.
 - Integra componentes aditivos y multiplicativos para mayor flexibilidad.
- **TBATS (Trigonometric, Box-Cox, ARMA, Trend, Seasonal)**
 - Extiende el modelo **ETS (Error-Trend-Seasonality)**.
 - Maneja **múltiples estacionalidades** y aplica **transformaciones Box-Cox**.

Modelos de Machine Learning: Aprovechan **múltiples variables predictoras y transformaciones** (lags, promedios móviles, calendarios).

- **LightGBM**
 - Modelo de *boosting* basado en árboles de decisión.
 - Alta eficiencia, escalabilidad y regularización.
- **XGBoost**
 - Algoritmo de *boosting* robusto y de alto rendimiento.
 - Amplio uso en problemas de regresión y pronóstico.
- **Ridge**
 - Regresión lineal con regularización L2.
 - Reduce la varianza y mejora la **generalización del modelo**.

Manejo de estacionalidad

1. Modelos Estadísticos

En los modelos estadísticos, la estacionalidad se maneja de forma interna a través de los parámetros que cada metodología selecciona automáticamente:

- ARIMA determina los parámetros estacionales (P, D, Q) y la necesidad de diferenciación estacional cuando corresponde.
- TBATS modela estacionalidades simples y múltiples mediante componentes trigonométricos y descomposiciones complejas.
- CES incorpora estacionalidad dentro de su estructura de suavizamiento exponencial, sin necesidad de transformar previamente la serie.

Gracias a este diseño, no se requiere desestacionalizar ni transformar manualmente la serie para capturar patrones estacionales.

2. Modelos de Machine Learning

No es obligatorio eliminar la estacionalidad en modelos de Machine Learning.

Lo indispensable es representarla en las features, ya que estos modelos solo pueden aprenderla si está explícitamente codificada en los datos.

A diferencia de los modelos estadísticos, los algoritmos de Machine Learning no reconocen estacionalidad ni dependencia temporal de forma automática. Por ello, se aplicaron transformaciones y *features* específicas que permiten representar estos patrones de manera explícita.

Para mejorar la estabilidad del target se aplicó una diferencia de primer orden, útil para reducir la tendencia y facilitar el aprendizaje del modelo. No obstante, esta transformación no elimina la estacionalidad, por lo que se incorporó información temporal adicional mediante *features*.

Métricas de desempeño

En un entorno hospitalario, donde cada centro presenta volúmenes y patrones de consumo heterogéneos, resulta fundamental emplear métricas que permitan comparar el desempeño de los modelos de manera justa entre distintas combinaciones hospital–producto. Para ello, se utilizaron métricas escaladas y porcentuales, las cuales normalizan el error y permiten interpretarlo en términos relativos y comparables.

En particular, para el cálculo del MASE se utilizó un modelo naïve estacional de 12 meses, que asume que la demanda futura será igual a la observada en el mismo mes del año anterior. Este enfoque constituye un benchmark ampliamente aceptado en la literatura, debido a su simplicidad, interpretabilidad y capacidad para capturar estacionalidades anuales típicas en series de demanda de insumos clínicos. Escalar los errores respecto a este modelo base permite evaluar si los modelos propuestos efectivamente mejoran la predicción sobre una referencia mínima pero robusta.

Las métricas utilizadas fueron:

- MASE (Mean Absolute Scaled Error): mide el error absoluto promedio del modelo en relación con el pronóstico naïve, facilitando comparaciones entre series con diferentes magnitudes.
- sMAPE (Symmetric Mean Absolute Percentage Error): cuantifica el error porcentual medio de forma simétrica, entregando una interpretación relativa independiente de la escala.

Estas métricas permiten evaluar el rendimiento predictivo de manera consistente y alineada con las mejores prácticas en la literatura de pronósticos.

Métrica	Valor	Interpretación
MASE	< 0.5	Desempeño excelente: el modelo presenta errores muy bajos respecto al pronóstico naïve. Ideal en entornos con alta estabilidad o estacionalidad bien capturada.
MASE	0.5 – 0.8	Rango típico de modelos buenos en la práctica.
MASE	0.8 – 1.0	Aceptable incluso para datos con alta variabilidad.
MASE	> 1.0	El modelo es peor que el naïve — necesita revisión.

Métrica	Valor	Interpretación
---------	-------	----------------

sMAPE	< 10 %	Excelente desempeño
sMAPE	~10 % - 20 %	Buen desempeño, realista en entornos variables
sMAPE	~20 % - 35 %	Moderado, puede ser aceptable según la variabilidad
sMAPE	> 35 %	Precisión baja, revisar modelo o datos

Optimización de hiperparámetros

La optimización de hiperparámetros se realizó de manera individual para cada combinación hospital–producto, siguiendo la misma lógica adoptada en el modelamiento. Dado que cada serie presenta dinámicas temporales y estructuras distintas, optimizar los modelos de forma independiente permite obtener configuraciones más adecuadas a la variabilidad local de cada hospital y producto, evitando soluciones globales subóptimas y mejorando la capacidad predictiva en contextos heterogéneos.

Modelos estadísticos.

Para la optimización de los **modelos estadísticos**, se utilizaron las versiones **Auto** de la librería **Nixtla**, las cuales permiten determinar automáticamente la mejor configuración de parámetros para cada método. Estos procedimientos exploran diversas combinaciones de parámetros relevantes —como tendencia, estacionalidad y factores de suavizamiento— y ajustan cada configuración a los datos históricos.

El desempeño de cada modelo candidato se evalúa mediante el **Criterio de Información de Akaike (AIC)**, un indicador que balancea precisión y complejidad. Finalmente, se selecciona la configuración con el **AIC más bajo**, correspondiente al modelo que ofrece el mejor equilibrio entre capacidad predictiva y parsimonia.

Modelos de machine learning.

En los modelos de Machine Learning, la optimización de hiperparámetros se llevó a cabo utilizando Optuna, un framework de búsqueda bayesiana eficiente para explorar espacios complejos. Para cada uno de los algoritmos —LightGBM, XGBoost y Ridge Regression— se definieron espacios de búsqueda específicos de acuerdo con sus arquitecturas particulares y el rol de cada parámetro en la regularización, profundización del árbol o estabilidad del modelo. Asimismo, para garantizar una exploración amplia y consistente, se ejecutaron 100 trials por cada combinación hospital–producto, permitiendo identificar configuraciones robustas y adaptadas a las características temporales de cada serie.

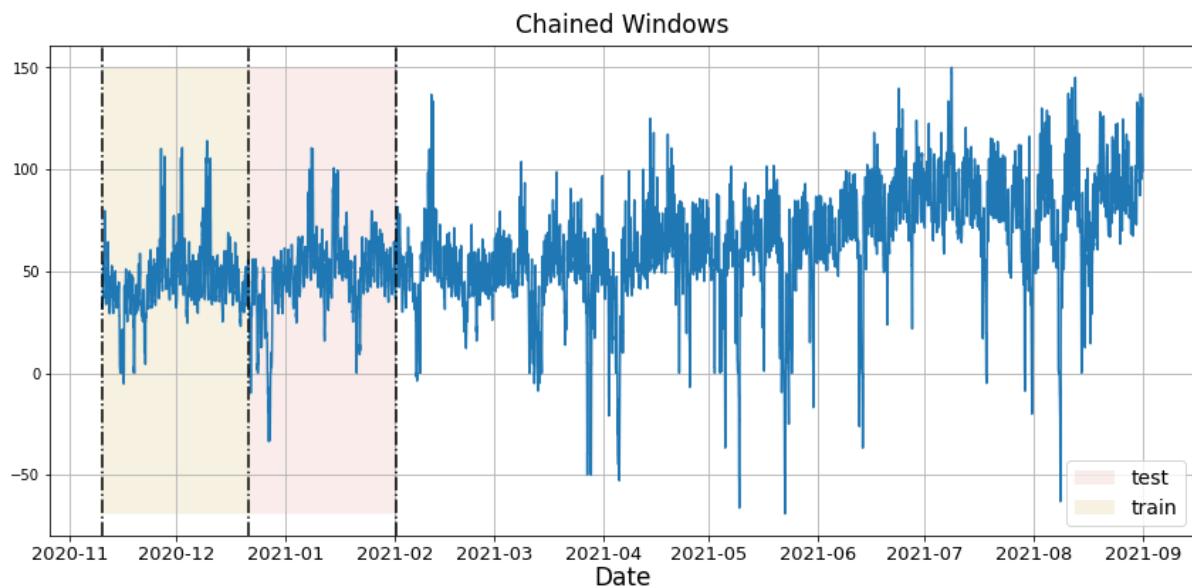
A continuación, se presenta un resumen de los principales hiperparámetros optimizados para cada algoritmo y su propósito dentro del proceso de modelado:

Parámetro	LightGBM	XGBoost	Ridge	Descripción general
Número de árboles (n_estimators)	✓	✓	—	Controla la cantidad de iteraciones o árboles del modelo.
Profundidad máxima (max_depth)	—	✓	—	Define la complejidad de los árboles para evitar sobreajuste.
Tasa de aprendizaje (learning_rate)	✓	✓	—	Ajusta la velocidad con que el modelo aprende en cada iteración.
Número de hojas (num_leaves)	✓	—	—	Regula la complejidad de los árboles en LightGBM.
Submuestreo de datos (bagging_fraction / subsample)	✓	✓	—	Usa una fracción de los datos para cada árbol, mejorando la generalización.
Submuestreo de variables (feature_fraction / colsample_bytree)	✓	✓	—	Usa una fracción de las variables para construir cada árbol.
Regularización L1 (lambda_l1 / reg_alpha)	✓	✓	—	Penaliza pesos grandes, favoreciendo modelos más simples.
Regularización L2 (lambda_l2 / reg_lambda / alpha)	✓	✓	✓	Reduce la varianza del modelo y previene el sobreajuste.
Peso mínimo por hoja (min_child_weight)	—	✓	—	Evita divisiones en nodos con poca información.
Intercepto (fit_intercept)	—	—	✓	Determina si se ajusta un término constante en la regresión.
Función objetivo (objective)	✓	—	—	Define el tipo de error a minimizar (por ejemplo, L1 o L2).

Este proceso de optimización sistemática permitió seleccionar, para cada modelo, la combinación de hiperparámetros que maximiza el desempeño predictivo respetando la estructura temporal de los datos y reduciendo el riesgo de sobreajuste.

Evaluación: Cross Validation

Para la evaluación de los modelos de predicción se aplicó un esquema de validación cruzada temporal con ventanas de prueba de 3 meses, acorde a las necesidades del negocio, que requiere planificar sus acciones comerciales con dicho horizonte. En cada iteración, el modelo se entrenó únicamente con información histórica y se evaluó sobre períodos futuros, simulando su desempeño real en producción. Este enfoque aumenta la confiabilidad de las predicciones, evita el sobreajuste y mejora la robustez del pronóstico al validar el comportamiento del modelo en múltiples ventanas temporales.



Resultados

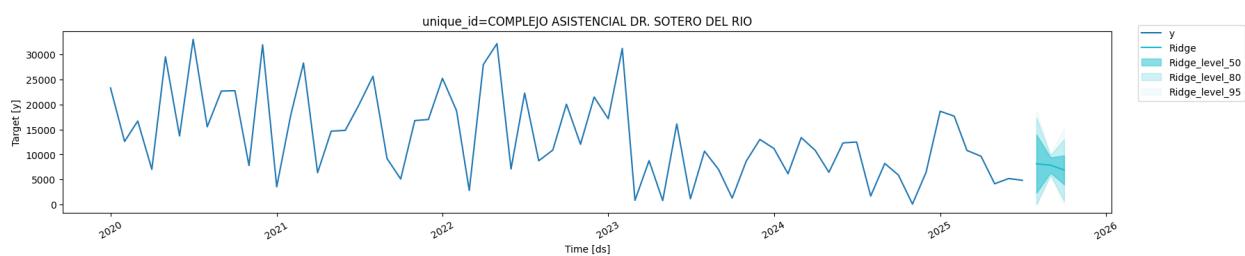
A continuación, se presentan los resultados obtenidos para cada combinación de producto y hospital, considerando los distintos enfoques utilizados: modelos estadísticos (ARIMA, CES, TBATS) y modelos de machine learning (LightGBM, XGBoost y Ridge).

Los modelos fueron evaluados mediante un esquema de validación cruzada temporal con ventanas de 3 meses, utilizando métricas de error ampliamente empleadas en series de tiempo, principalmente sMAPE y MASE, que permiten comparar el desempeño entre hospitales y productos de manera consistente.

Los siguientes cuadros resumen el rendimiento alcanzado por cada modelo para los insumos seleccionados — Apósitos y Guantes Médicos — en los hospitales Sótero del Río y El Salvador, permitiendo identificar qué enfoques entregan predicciones más precisas en cada contexto operativo.

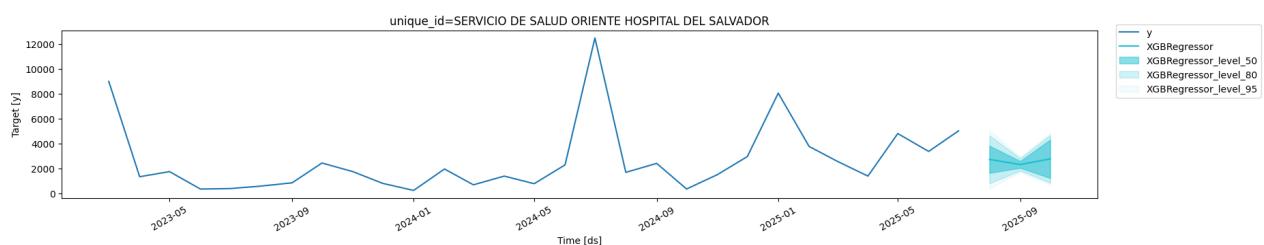
Resultados: Apósitos - Sótero del Río

modelo	smape	mase
ARIMA	34.52%	0.5958
TBATS	39.56%	0.7016
CES	34.84%	0.6565
LGBMRegressor	35.25%	0.6885
Ridge	32.07%	0.5413
XGBRegressor	40.56%	0.8945



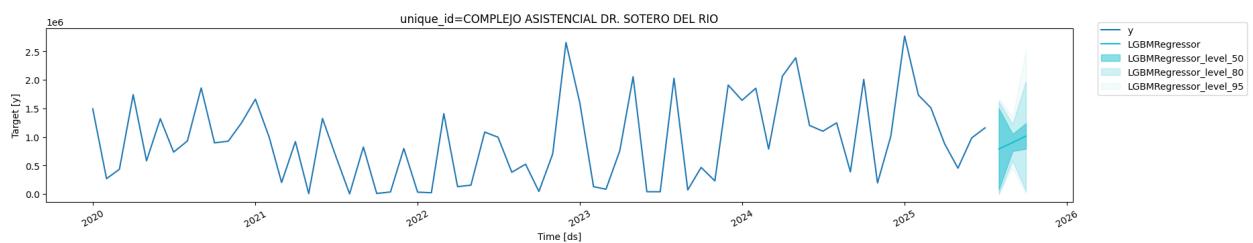
Resultados: Apósitos - El Salvador

modelo	smape	mase
ARIMA	24.92%	0.5429
TBATS	50.03%	0.8874
CES	31.56%	0.6446
LGBMRegressor	24.40%	0.5345
Ridge	25.75%	0.5761
XGBRegressor	22.66%	0.4974



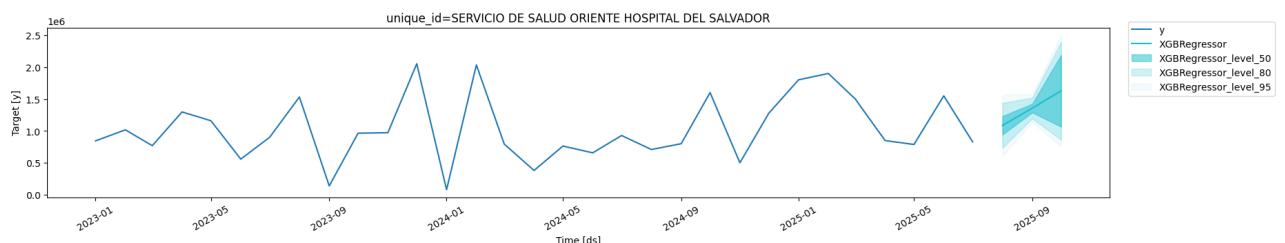
Resultados: Guantes Médicos - Sótero del Río

modelo	smape	mase
ARIMA	29.15%	0.8699
TBATS	27.12%	0.7957
CES	29.43%	0.9095
LGBMRegressor	23.93%	0.6860
Ridge	27.76%	0.8628
XGBRegressor	24.73%	0.7370



Resultados: Guantes Médicos - El Salvador

modelo	smape	mase
ARIMA	21.64%	0.8767
TBATS	19.69%	0.7973
CES	20.91%	0.8084
LGBMRegressor	17.25%	0.7830
Ridge	23.13%	1.0247
XGBRegressor	13.78%	0.5700



Análisis de resultados

- Modelos clásicos (ARIMA, TBATS, CES)
 - Resultan adecuados cuando la demanda presenta **patrones estables** y poco ruido.
 - En hospitales donde el consumo es **más predecible**, estos modelos siguen siendo una referencia.
- Modelos de Machine Learning
 - **XGBoost y LightGBM** destacan cuando hay irregularidades, fluctuaciones marcadas o patrones más complejos, ya que sus métodos de *boosting* permiten capturar relaciones no lineales y efectos de interacción.
 - **Ridge** resulta más apropiado en series estables y con variaciones suaves, donde las relaciones lineales explican bien la dinámica de la demanda.
- Los mejores modelos alcanzaron **sMAPE <= 32.07%** y **MASE <= 0.69**

producto	hospital	modelo	smape	mase
Apósito	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Ridge	32.07%	0.5413
Apósito	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	XGBRegressor	22.66%	0.4974
Guantes	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	LGBMRegressor	23.93%	0.6860
Guantes	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	XGBRegressor	13.78%	0.5700

Etapa 6: Explotación del modelo y recomendación

En esta última fase del pipeline, se aplica el modelo entrenado para estimar la demanda futura de productos médicos por hospital, con fines de planificación operativa y priorización comercial.

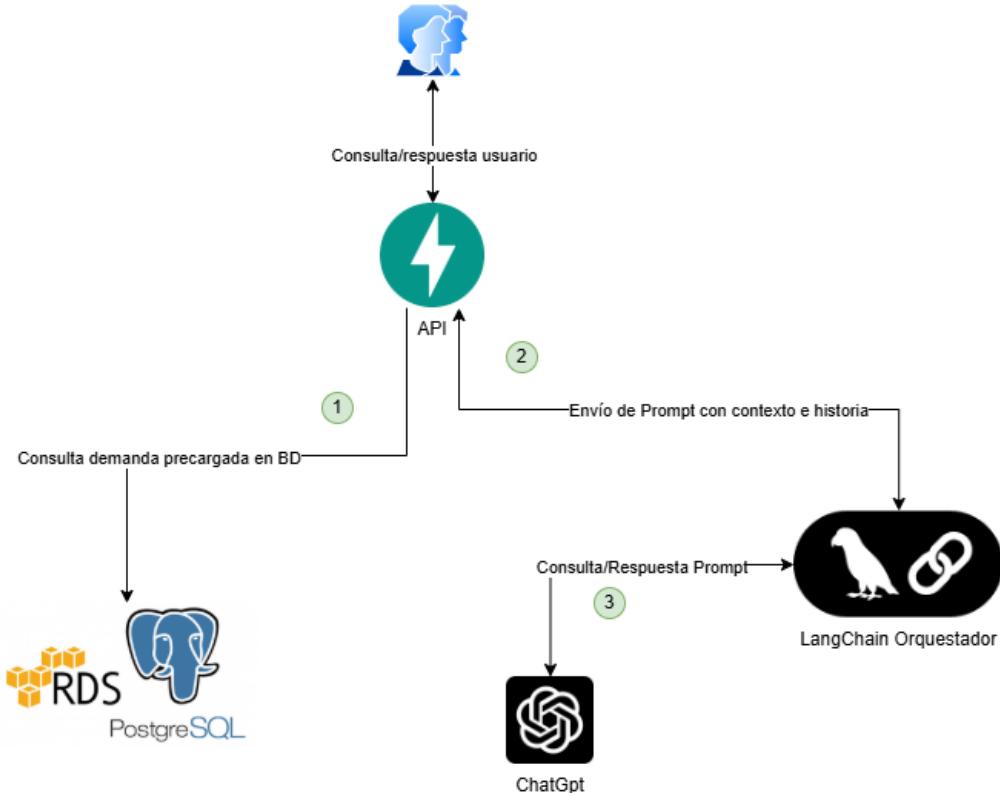
Actividades a realizar

Generación de predicciones futuras: se estima la cantidad demandada para el siguiente periodo mensual ([año_mes + 1](#)) por cada combinación hospital-producto.

- **Exportación de resultados:** los valores predichos se consolidan en una tabla de salida estructurada, que incluye los identificadores del hospital, producto y periodo, junto con la demanda proyectada.
- **Construcción de un ranking de prioridades comerciales:** se ordenan las combinaciones hospital-producto según el volumen de demanda esperada, permitiendo identificar oportunidades de alto impacto para acciones comerciales, planificación logística o visitas de representantes.

Este proceso permite a la empresa anticipar comportamientos de compra y focalizar recursos de manera estratégica en función de la demanda proyectada.

Finalmente los resultados serán la base para una API a desarrollar, la cual usará esta fuente para presentar respuestas a consultas sobre demandas de hospitales, la arquitectura propuesta es la siguiente:



El uso de la API (en Postman) sería como lo siguiente:



```

1  {
2    "session_id": "111555222",
3    "question": "Que hospital me recomiendas visitar y que producto ofrezco para el
4    próximo mes?"
  }
  
```

200 OK

Body JSON Preview Visualize

```

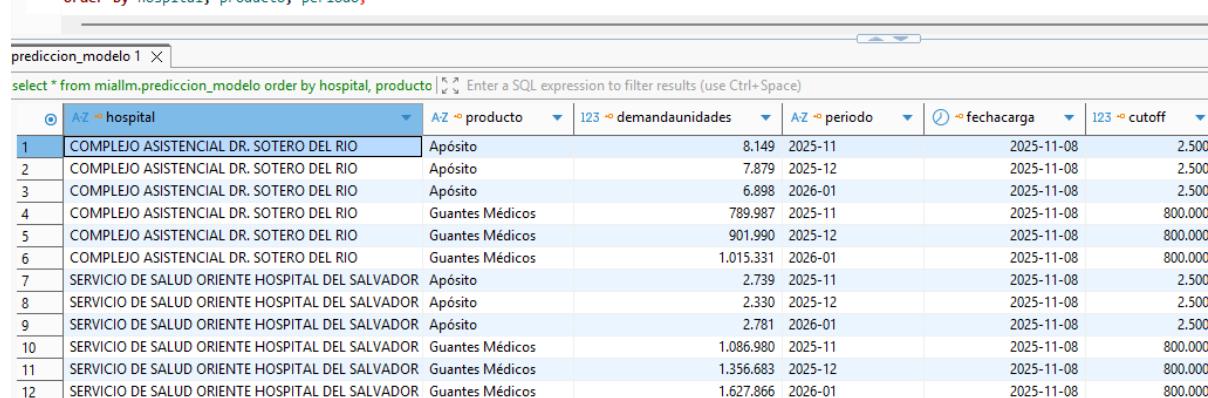
1
2   "answer": "Te recomiendo visitar el Hospital Clínico Metropolitano El Carmen
3   Doctor Luis Valentín Ferrada y ofrecer el apósito Tegaderm transparente
4   adhesivo de 44 x 44 cm, 100 unidades, ya que tienen una demanda prevista
5   de 20,000 unidades. Para eficientar tu capacidad de generar demanda,
6   prepara una presentación destacando la calidad y beneficios del producto,
7   y ofrece muestras gratuitas para que puedan evaluar su efectividad."
  
```

Validación funcionamiento correcto

Para corroborar que los resultados de la API son correctos, dado que está basado en una GenAI, la cual, de no gestionar correctamente los Prompt Template y la información que se le proporciona es que se evaluaron una serie de Test generales para corroborar su resultado. Las pruebas no son más que preguntas a la API, las cuales posteriormente fueron validadas manualmente. Estas son las siguientes:

- **Cual es la demanda de [PRODUCTO] para el mes de [PERIODO] en [HOSPITAL]?**
 - Propósito: Acá lo que se busca es confirmar que la API entregue respuestas exactas y reales basado en la información que le fue entregada desde la Base de datos con las predicciones calculadas. La pregunta indicada acá es una generalidad de todas las que se hicieron y deben coincidir con la información de la base de datos de predicciones. El cual se refleja en la siguiente imagen que representa una consulta a la tabla de la base de datos con las predicciones obtenidas.

select * from miallm.prediccion_modelo
order by hospital, producto;



The table displays the following data:

	AZ ~ hospital	AZ ~ producto	123 ~ demandaunidades	AZ ~ periodo	0 ~ fechacarga	123 ~ cutoff
1	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Apósito	8.149	2025-11	2025-11-08	2.500
2	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Apósito	7.879	2025-12	2025-11-08	2.500
3	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Apósito	6.898	2026-01	2025-11-08	2.500
4	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Guantes Médicos	789.987	2025-11	2025-11-08	800.000
5	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Guantes Médicos	901.990	2025-12	2025-11-08	800.000
6	COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO	Guantes Médicos	1.015.331	2026-01	2025-11-08	800.000
7	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Apósito	2.739	2025-11	2025-11-08	2.500
8	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Apósito	2.330	2025-12	2025-11-08	2.500
9	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Apósito	2.781	2026-01	2025-11-08	2.500
10	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Guantes Médicos	1.086.980	2025-11	2025-11-08	800.000
11	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Guantes Médicos	1.356.683	2025-12	2025-11-08	800.000
12	SERVICIO DE SALUD ORIENTE HOSPITAL DEL SALVADOR	Guantes Médicos	1.627.866	2026-01	2025-11-08	800.000

- Respuestas para [PRODUCTO]=Apósitos, [PERIODO]=Diciembre y [HOSPITAL]= Sótero del Río: La demanda prevista de apósitos para el Complejo Asistencial Dr. Sótero del Río en diciembre de 2025 es de 7879 unidades, superando el cut-off de 2500 unidades. Te recomiendo priorizar este hospital para la venta de apósitos durante ese mes, ya que es un objetivo rentable. Asegúrate de preparar una propuesta competitiva y resaltar la calidad y disponibilidad inmediata del producto para maximizar las posibilidades de éxito en la venta.
- Respuesta para el resto de los registros de las predicciones:

Hospital	Producto	Periodo	Demand en	Demand en
----------	----------	---------	-----------	-----------

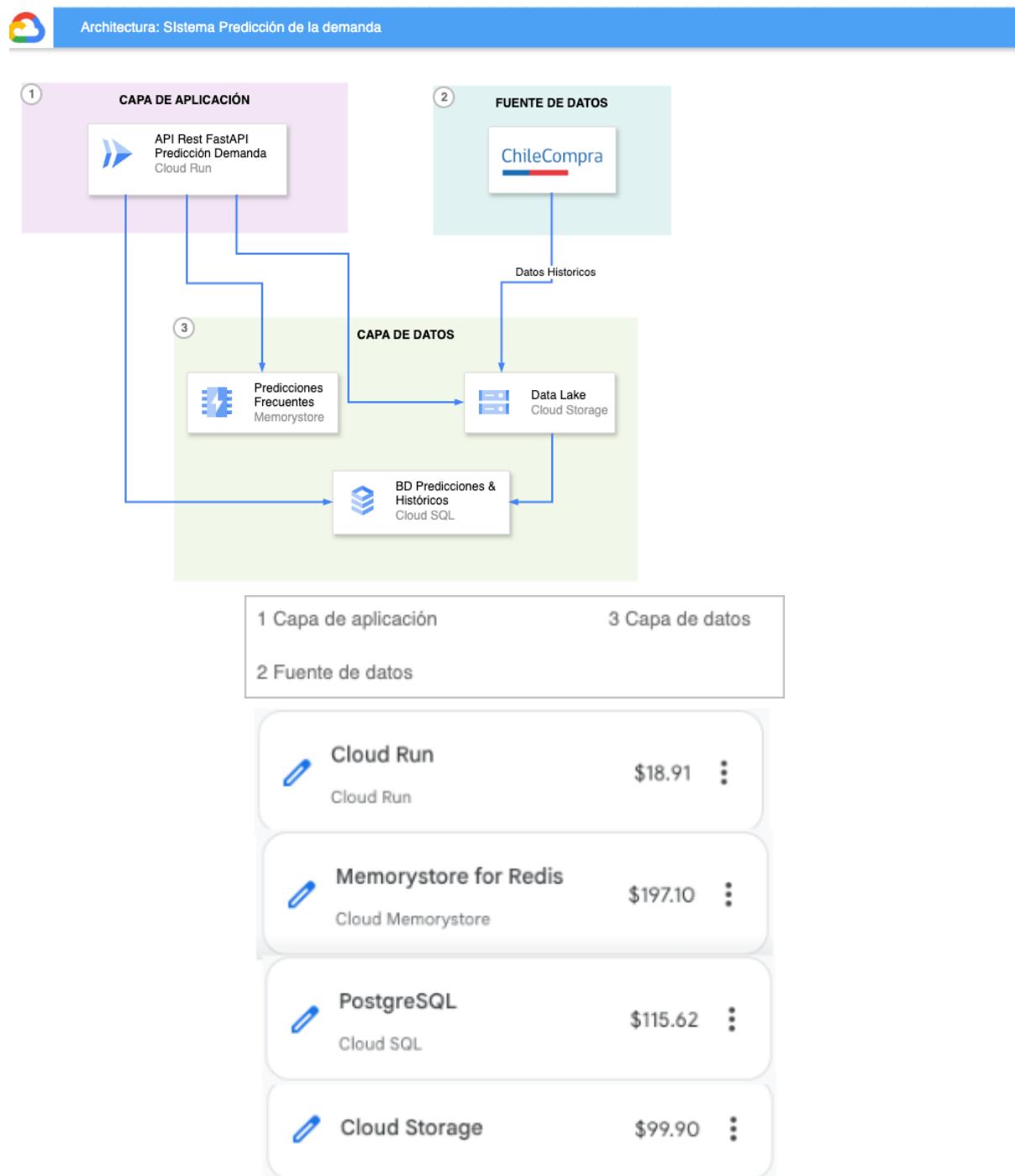
			respuesta API	Base de datos
Sótero del Río	Apósitos	Noviembre	8.149	8.149
Sótero del Río	Apósitos	Diciembre	7.879	7.879
Sótero del Río	Apósitos	Enero	6.898	6.898
Sótero del Río	Guantes Médicos	Noviembre	789.987	789.987
Sótero del Río	Guantes Médicos	Diciembre	901.990	901.990
Sótero del Río	Guantes Médicos	Enero	1.015.331	1.015.331
Hospital El Salvador	Apósitos	Noviembre	2.739	2.739
Hospital El Salvador	Apósitos	Diciembre	2.330	2.330
Hospital El Salvador	Apósitos	Enero	2.781	2.781
Hospital El Salvador	Guantes Médicos	Noviembre	1.086.980	1.086.980
Hospital El Salvador	Guantes Médicos	Diciembre	1.356.683	1.356.683
Hospital El Salvador	Guantes Médicos	Enero	1.627.866	1.627.866

- **Cual es la demanda de Apositos para el mes de Febrero en el hospital Sótero del Río?**
 - **Propósito:** Tanto los Apósitos como el hospital Sótero del Río tienen predicciones en la base de datos, sin embargo el periodo que se menciona acá es inexistente, por lo tanto la respuesta debe indicar aquello.
 - **Respuesta:** Lo siento, no tengo información sobre la demanda de apósitos para febrero de 2026 en el Complejo Asistencial Dr. Sótero del Río. Te recomiendo centrar tus esfuerzos en los meses y productos donde tenemos datos disponibles. Por ejemplo, prioriza la venta de apósitos en diciembre de 2025 en el mismo hospital, donde la demanda prevista es de 7879 unidades, superando el cut-off, lo que lo hace un objetivo rentable. Asegúrate de destacar la calidad del producto y ofrecer condiciones atractivas para cerrar la venta.
- **Cual es la demanda de Apósitos para el mes de Febrero en Integramedica?**

- **Propósito:** En este caso los Apósitos y el Periodo si existe en la base de datos, sin embargo el recinto de salud no lo está. Por lo tanto debe entregar una respuesta acorde a esa situación:
- **Respuesta:** Lo siento, no tengo información sobre la demanda de apósitos para febrero de 2026 en Integramedica. Te recomiendo centrar tus esfuerzos en los hospitales del Sistema de Salud Público de Chile donde tenemos datos disponibles. Por ejemplo, prioriza la venta de apósitos en el Complejo Asistencial Dr. Sótero del Río para diciembre de 2025, donde la demanda prevista es de 7879 unidades, superando el cut-off. Asegúrate de destacar la calidad del producto y ofrecer condiciones atractivas para cerrar la venta
- **Entregame la demanda de cintas médicas para el mes de Febrero del 2026 en la clínica Integramédica?**
 - **Propósito:** Validar que la respuesta que entregue sea directamente relacionada con los datos de las predicción previamente calculados, por lo que solo puede entregar la demanda de productos para hospitales que sí fueron hechas, y en este caso, ni la cinta médica ni Integramedica es parte de esta predicción.
 - **Respuesta:** Lo siento, no tengo información sobre la demanda de cintas médicas para febrero de 2026 en la Clínica Integramedica. Te recomiendo centrar tus esfuerzos en los productos y hospitales del Sistema de Salud Público de Chile donde tenemos datos disponibles. Por ejemplo, enfocar tus visitas en el Servicio de Salud Oriente Hospital del Salvador para guantes médicos en enero de 2026, donde la demanda prevista es alta y supera el cut-off.
- **Que criptomoneda me recomiendas para invertir?**
 - **Propósito:** Verificar que la respuesta solo se aadecue al contexto que deseamos, en este caso, la venta de insumos médicos y la predicción de la demanda de estos. En caso de que sea una pregunta no relacionada a esto la API debe ser capaz de responder no solo indicando que no tiene información al respecto de aquello, si no que además proponiendo al usuario que realice preguntas para el contexto que deseamos.
 - **Respuesta:** Lo siento, no tengo información sobre criptomonedas. Te recomiendo centrar tus esfuerzos en el negocio de la salud, donde podemos asesorarte mejor. Por ejemplo, enfoca tus visitas en el Complejo Asistencial Dr. Sótero del Río para guantes médicos en diciembre de 2025, ya que la demanda prevista supera el cut-off, lo que lo hace un objetivo rentable y prioritario.

7. Análisis de sensibilidad y costos

Arquitectura



Supuestos

1. Situación actual

- En promedio, los analistas dedican *8 horas semanales* a la identificación manual de oportunidades de negocio en el sistema de salud público.
- Esto representa *32 horas mensuales* por analista y un costo asociado de entre **\$244.380** y **\$716.365** pesos *mensuales* considerando el valor/hora del recurso (según <https://tusalario.org/>).

2. Proyección con IA

- La solución basada en inteligencia artificial automatiza la búsqueda y filtrado de oportunidades, reduciendo el tiempo de análisis en un **80%**.
- El tiempo requerido pasaría de *8 horas mensuales* a solo *1.6 hora*.

3. Impacto económico

- **Ahorro de tiempo:** Reducción de *25.6 horas mensuales* por analista → equivalente a entre **\$2.346.048** y **\$6.877.100** anuales por profesional en el equipo.
- **Mayor productividad:** El tiempo liberado se puede destinar a actividades de mayor valor (ejemplo: análisis estratégico, cierre de acuerdos).

Cut-Off

Alineación entre la decisión analítica y la estrategia comercial

Objetivo	Qué busca	En el caso práctico
Eficiencia operativa	Reducir esfuerzos en oportunidades poco rentables	Solo visitar hospitales donde el modelo predice alta cobertura
Rentabilidad económica	Maximizar retorno sobre costo de visita / acción	Balancear costo de acción (visita comercial) vs. potencial de cobertura

Control del riesgo	Evitar sobre-predicción	Fijar umbral mínimo de confianza en la predicción
---------------------------	-------------------------	---

Mínimo de unidades esperadas que justifican una acción comercial

Hospital	Producto	Cut-Off
Sótero del Río	Apósitos	2.500 Unidades
Sótero del Río	Guantes Médicos	800.000 Unidades
El Salvador	Apósitos	2.500 Unidades
El Salvador	Guantes Médicos	800.000 Unidades

Analisis Costos

Recordemos los supuestos económicos para los costos por factor humano:

Situación actual:

- 8 horas semanales dedicadas al análisis manual (32 horas mensuales).
- Costo mensual por analista: \$244.380 a \$716.365 CLP.
- Costo anual por analista: \$2.932.560 a \$8.596.380 CLP.

Con modelo IA:

- Reducción del 80% en tiempo de análisis.
- Ahorro mensual: \$195.504 a \$573.092 CLP.
- Ahorro anual: \$2.346.048 a \$6.877.100 CLP por analista.

Costos por implementación de tecnología en la nube:

Año	Concepto	Costo (CLP)	Factor de descuento (1/(1+r)^t)	Valor presente (CLP)
0	Desarrollo e integración	5.000.000	1.000	5.000.000
1	Operación (Despliegue + nube + soporte)	1.500.000	0.909	1.364.000
2	Operación (Despliegue + nube + soporte)	1.500.000	0.826	1.239.000
3	Operación (Despliegue + nube + soporte)	1.500.000	0.751	1.126.000

$$\mathbf{NPV = 5.000.000 + 1.364.000 + 1.239.000 + 1.126.000 = 8.729.000 CLP}$$

ROI y Payback revisados

Escenario	Ahorro Anual Total (5 analistas)	Ahorro 3 Años	Costos Totales	Beneficio Neto (NPV aprox.)
Pesimista	\$11.730.240	\$35.190.720	\$8.729.000	+\$26.461.720
Base	\$22.565.266	\$67.965.780	\$8.729.000	+\$59.236.780
Optimista	\$34.385.500	\$103.156.500	\$8.729.000	+\$94.427.500

Escenario	Fórmula ROI	ROI (%)	Fórmula Payback	Payback
Pesimista	$(26.46M / 8.73M) \times 100$	≈ 303 %	8.73M / 11.73M	0.74 años (≈ 9 meses)
Base	$(59.24M / 8.73M) \times 100$	≈ 678 %	8.73M / 22.56M	0.38 años (≈ 4.6 meses)
Optimista	$(94.43M / 8.73M) \times 100$	≈ 1082 %	8.73M / 34.38M	0.25 años (≈ 3 meses)

El análisis económico muestra que la implementación de la solución IA en Solventum es altamente rentable.

Con una inversión total de \$8,7 millones, el proyecto se paga en menos de 5 meses y genera un ROI estimado superior al 600% en tres años.

Incluso en escenarios conservadores, la solución sigue aportando valor económico y operacional.

8. Conclusiones y trabajos futuros

¿Se lograron los objetivos?

<u>Objetivo</u>	<u>Resultado</u>	<u>Status</u>
Estandarizar y estructurar las descripciones de los ítems de las órdenes de compra , asociándolos correctamente con los productos del catálogo de Solventum, logrando una tasa de correspondencia mínima del 90%.	Se genera un identificador único para los productos de Solventum seleccionados (Apósitos y Guantes) de tal manera que permite identificar en la base histórica a los mismos y permitir usarlos para el análisis por parte de los vendedores de Solventum como para el modelo de predicción. Con una correspondencia del 95%.	✓
Desarrollar un modelo predictivo supervisado que estime la demanda potencial de productos en cada hospital , con un error relativo medio (MASE) menor a 0.7 y un SMAPE menor a 35%.	Los modelos supervisados desarrollados (Ridge, XGBoost, LightGBM) alcanzaron un MASE inferior a 0.686 y SMAPE menor al 33%	✓
Diseñar una herramienta asistida por GenAI que interprete el forecast y sugiera acciones comerciales , permitiendo a los equipos priorizar hospitales de forma ágil y reduciendo el tiempo de análisis en al menos un 20% frente al proceso actual.	Se construye un API que implementa un modelo de GenAI que permite consultas de ventas en lenguaje natural y que responde recomendaciones basado en el resultado del modelo predictivo.	✓

Entregables

- Base histórica de Órdenes de Compra (Data raw).
- Base histórica de OC con datos preprocesados y generados (Cantidad Real, Precio Real, Etiquetado/Estandarización productos catálogo).
- Modelo predictivo de la demanda de categoría de productos por hospital.
- API con modelo generativo de recomendación que considera la demanda de los hospitales.

Trabajos futuros

- **Escalar el Alcance**
 - Incorporar más hospitales en el análisis
 - Evaluar más categorías y productos del catálogo Solventum
 - Priorizar oportunidades por región y tamaño del hospital
- **Mejoras en Calidad de Datos**
 - Estandarizar descripciones de productos
 - Unificar unidades de medida y codificación
 - Reducir ruido en registros históricos de órdenes de compra
- **Incorporación de Información Externa**
 - Estacionalidad clínica (p. ej., ciclos quirúrgicos)
 - Indicadores operativos hospitalarios (ocupación, carga UCI / Qx)
 - Cambios en políticas de abastecimiento público
- **Mejoras en Modelado y Técnica**
 - Ajuste y reentrenamiento continuo del forecast
 - Pipeline mensual automatizado (datos → forecast → tablero dinámico)
 - Alertas ante cambios abruptos en demanda proyectada
- **Medición del Impacto**
 - Comparación antes vs después de la adopción del modelo
 - KPIs sugeridos:
 - Abastecimiento por producto o categoría

- Activación / reactivación de hospitales
- Conversión visitas → órdenes de compra

Trabajos futuros - Recomendación compañía

Gestión del Cambio

- **Crear un Centro de Excelencia en Analítica (AI CoE)**
 - Equipo responsable de mantener modelos, datos y metodologías
 - Rol articulador entre áreas comerciales, TI y abastecimiento
 - Asegura continuidad y evolución del proyecto en el tiempo
- **Capacitación y Desarrollo de Capacidades Internas**
 - Formación en analítica avanzada y gobierno de datos para equipos clave
 - Talleres prácticos orientados al uso del forecast y herramientas de soporte
 - Manual de adopción y guía operativa para el equipo comercial
- **Integración del Modelo en la Operación Comercial**
 - Definir responsables, periodicidad y flujo de actualización del forecast
 - Incorporar insights en reuniones de planificación comercial y visitas a hospitales
 - Indicadores claros de seguimiento y retroalimentación continua
- **Cultura Organizacional Basada en Evidencia**
 - Decisiones sustentadas en datos antes que criterios subjetivos
 - Mayor consistencia y velocidad en la ejecución comercial
 - Promueve transparencia, trazabilidad y mejora continua

Anexo

Diccionario de Datos - Documentos de compras de productos médicos

Nombre de Columna	Tipo de Dato	Descripción
codigoOC	object	Identificador o código único de la orden.
FechaEnvioOC	object	Fecha relacionada con la orden de compra.
NombreOC	object	Nombre relacionado con la orden o proveedor.
DescripcionOC	object	Descripción detallada del contenido o propósito.
EstadoOC	object	Estado actual de la orden de compra.
ProcedenciaOC	object	Ejemplos: Trato Directo
MonedaOC	object	Ejemplos: CLP, USD, CLF
MontoNetoOC	object	Monto o valor económico de la orden.
DescuentosOC	int64	Ejemplos: 0
CargosOC	int64	Ejemplos: 0, 250000
ImpuestosOC	object	Ejemplos: 1140000, 427500, 349600, 893235,41, 0
MontoTotalOC	object	Monto o valor económico de la orden.
ImpuestosOC_CLP	object	Ejemplos: 1140000, 427500, 349600, 893235,41, 0
MontoNetoOC_CLP	object	Monto o valor económico de la orden.

MetodoPago	object	Ejemplos: 30 dias contra la recepcion conforme de la factura
TipoDespacho	object	Ejemplos: Despachar a Dirección de envío, Otra Forma de Despacho, A convenir
Financiamiento	float64	Ejemplos: nan, 220400400102.0
UnidadCompra	object	Unidad organizacional o región asociada.
UnidadCompraRUT	object	Rol Único Tributario, identificador fiscal en Chile.
RegionUnidadCompra	object	Unidad organizacional o región asociada.
entCode	int64	Ejemplos: 7320
Institucion	object	Ejemplos: COMPLEJO ASISTENCIAL DR. SOTERO DEL RIO
Sector	object	Ejemplos: SALUD
Proveedor	object	Información del proveedor relacionado.
ProveedorRUT	object	Información del proveedor relacionado.
ActividadProveedor	object	Información del proveedor relacionado.
TamanoProveedor	object	Información del proveedor relacionado.
RegionProveedor	object	Información del proveedor relacionado.

RubroN1	object	Ejemplos: Equipamiento y suministros médicos, Servicios de producción y fabricación industrial, Salud, servicios sanitarios y alimentación, Medicamentos y productos farmacéuticos, Equipamiento para laboratorios
RubroN2	object	Ejemplos: Productos para imágenes y de medicina nuclear, Productos administrados por vía intravenosa y arterial, Servicios de apoyo o soporte a la fabricación, Productos quirúrgicos, Medicina y servicios de salud
RubroN3	object	Ejemplos: Marcapasos cardiacos y productos relacionados, Bombas, analizadores, sensores y accesorios de infusión intravenosa, Servicios de mantenimiento y reparación de equipo de fabricación, Equipo quirúrgico, accesorios y productos relacionados, Evaluaciones de salud individuales
CodigoProductoONU	int64	Identificador o código único de la orden.

Referencias bibliográficas.

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>
- Garza, F., Mergenthaler Canseco, M., Challú, C., & Olivares, K. G. (2022). StatsForecast: Lightning fast forecasting with statistical and econometric models. PyCon, Salt Lake City, Utah, United States. <https://github.com/Nixtla/statsforecast>